UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

APPLICATION OF POST-STACK SEISMIC ATTRIBUTES, WELL DYNAMIC DATA, AND MACHINE LEARNING FOR CARBONATE RESERVOIR FACIES AND PRODUCTIVITY PREDICTION FOR UPSTREAM OPTIMIZATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MARCUS VINICIUS RODRIGUES MAAS

Norman, Oklahoma

2025

APPLICATION OF POST-STACK SEISMIC ATTRIBUTES, WELL DYNAMIC DATA, AND MACHINE LEARNING FOR CARBONATE RESERVOIR FACIES AND PRODUCTIVITY PREDICTION FOR UPSTREAM OPTIMIZATION

A DISSERTATION APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Heather Bedle, Chair

Dr. Deepak Devegowda

Dr. Matthew J. Pranter

Dr. Gerilyn S. Soreghan

"Those who say it can't be done are usually interrupted by others doing it"

James Baldwin

Acknowledgements

First of all, thanks to the Almighty God of Abraham, Isaac and Jacob, and the Holy Family Jesus, St Mary, and St Joseph, who never abandoned me, ever.

To my beloved wife Elayne Cristina, who encouraged me to pursuit this PhD even after a long time being away from the Academy, and our beloved daughter Ellen, everything I did was to and for you! Thank you for your unconditional love and patience! I love you!

To my parents Werner Maas (*in memoriam*) and Nair Maas, who did not measure any effort to give me the best of education they could with limited resources. Thank you for making me believe in the power of education!

To PETROBRAS, the proud of the Brazilian People! For the job opportunity in the last 23 years and the total financial support of my doctoral research. I hope this dissertation can somehow contribute to its continuous improvement of efficiency!

To my academic advisor Dr. Heather Bedle, from the very beginning you trusted in my project and my realization capacity, giving me total freedom for creation and guiding me through this incredible journey of knowledge.

To Petrobras managers Jonilton Pessoa, Rogerio Cunha, Marcos Galvao, Henrique Penteado, Marcelo Matsdorf, and Andre Henrique for their invaluable support during my project approval, transference to the United States, and dataset releasing for my doctoral research.

To Petrobras colleagues and friends Conrado Keidel, Ricardo Tarabini, Joao Guerreiro, Marco Carlotto, Bruno Eustaquio, Cristiano Rancan, Caroline Cazarin (in memoriam), Eduardo Carvalho (in memoriam), Ilson Rubim, Alan Mossinger, Andre Spigolon, Glauco Braganca, and Ricardo Pereira for the help, ideas exchange and insightful information for my research development.

Many thanks to other AASPI researchers Dr. Kurt Marfurt (our eternal Xaman), Dr. David Lubo-Robles, Dr. Marcilio Matos, Mario Ricardo Ballinas, and all other AASPI and OU Geosciences colleagues.

Thanks to my committee members Dr. Matthew J. Pranter, Dr. Gerilyn S. Soreghan, and Dr. Deepak Devegowda for all of your knowledge and insightful classes. You contributed a lot to my academic formation and my research development.

To all OU Geosciences staff led by Rebecca Fay, your support is crucial for our academic goals achievement.

Last but not least, to the people of Oklahoma, an amazing place to live with very gentle people! I will never forget the time spent here!

Table of contents

Acknowledgements……………………………………………………………..……………..v

Table of contents………………………………………………………………..……….vii

List of figures……………………………………………………………...……….xii

List of tables…………………………………………………………………………xxxvii

Chapter 1 : Introduction…………………………………………………..………..1

References……………………………………………………………………...4

Chapter 2: Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data………..……..7

Abstract………………………………………………………………….………..7

Introduction………………………………………………………..……………..8

Area location and Geological Background……………………….…………………10

Dataset………………………………………………………..…………………15

Seismic data……………………………………………….……………….**.15**

Well data…………………………………………………………...…..**.18**

Method……………………………………………………...…………..19

Seismic data conditioning, interpretation and attribute extraction……………….**19**

Well data compilation…………………………………………………..………..21

Carbonate facies and related seismic attributes…………………………………23

Unsupervised clustering basic theory……………………………..………..27

Results……………………………………………………………………….…..29

Vertical sections……………………………………………………….…………30

3D view of the clustering models…………………………………………………31

Discussion…………………………………………………………..………..37

Conclusions……………………………………………………………...………39

Acknowledgements…………………………………………………….……39

References……………………………………………………………………….……40

Chapter 3- Unraveling the hidden relationships between seismic multi-attributes, well dynamic data, and Brazilian pre-salt carbonate reservoirs productivity: a shallow versus deep machine learning approach……………………………………………..…………..46

Abstract……………………………………………………………...……..46

Introduction…………………………………………………………...……...…50

Area location and Geological Background……………………………………………52

Dataset………………………………………………………………………...53

Seismic data……………………………………………………………...53

Well data………………………………………………………………………53

Method…………………………………………………………………………..54

Seismic data conditioning, and attribute extraction…………………………..…54

Well dynamic data compilation and training dataset structuration……………..56

Exploratory data analysis……………………………………………………..60

Basic concepts of selected supervised machine learning algorithms……………64

Random forest…………………………………………………………65

K-nearest neighbors…………………………………………………...66

Support vector machines………………………………………………67

Multi-layer perceptron………………………………………………...68

Data training, reservoir productivity prediction, and models validation………..70

Results……………………………………………………………………………...73

Supervised regression models…………………………………………………73

Vertical seismic sections………………………………………………………77

Discussion…………………………………………………………………………….81

Conclusions…………………………………………………………………………...85

Acknowledgements…………………………………………………………………..86

References……………………………………………………………………………87

Chapter 4- Transfer learning from seismic data for predicting reservoir productivity and lithologies for upstream optimization…………………………………………………………95

Abstract……………………………………………………………………………….95

Introduction…………………………………………………………………………..96

Area location and Geological Background…………………………………………98

Dataset……………………………………………………………………………….102

Method…………………………………………………………………………….106

Transfer learning basic aspects………………………………………………….107

Transfer learning feasibility studies…………………………………………….110

Principal component analysis (PCA) …………………………………………...110

Exploratory data analysis (EDA) ……………………………………………….112

Feature engineering for transfer learning optimization………………………...116

Results and discussions………………………………………………………..125

Bacalhau field application……………………………………………………...128

Lapa field application…………………………………………………………..132

Insights and implications……………………………………………………….135

Conclusions…………………………………………………………………….139

Acknowledgements……………………………………………………………140

References……………………………………………………………………...140

Chapter 5: Conclusions……………………………………………………..….150

Appendix 1 – State of art of petrography, petrogenesis and hydrothermal alteration of the Barra Velha Late Aptian lacustrine carbonates- Santos Basin, SE Brazil offshore………...154

Abstract………………………………………………………………………...154

Introduction…………………………………………………………………….155

Geological Background………………………………………………………..156

First proposed petrographic classification schemes and reservoir models …………159

From first pre-salt carbonate well penetrations to first commercial production and world class play emergency (2002-2014) **……………………………………………159**

The abiotic-biotic dichotomy (2014 – 2019) **……………………………………165**

The discovery of a regional scale pervasive hydrothermal alteration process (2017-2020) **…………………………………………………………………………...168**

Towards the unification theory and the public pre-Salt public data avaliability "spring" (2020 – 2024) **…………………………………………………………...175**

The relentless search for analogues **……………………………………………..179**

Discussion……………………………………………………………………...181

Conclusions……………………………………………………………………185

References……………………………………………………………………...186

Appendix 2 – Supervised regression toy examples…………………………………...193

Decision trees…………………………………………………………………..193

K-nearest neighbors (KNN) ……………………………………………………..197

Support vector machines………………………………………………………...200

Multilayer perceptron…………………………………………………………203

References……………………………………………………………………...207

List of Figures

Figure 2.1. Location map of the Mero Field and Central Libra Appraisal Plan…………..………11

Figure 2.2. Location map of the Mero Field and Central Libra Appraisal Plan with the wells and seismic dataset used in this study. Wells A to H have drill stem test (DST) data. Wells X1 to X5 were drilled for appraisal purposes. Wells A and C were selected as an injector-producer pair to support an anticipated production system test…………………………………………………11

Figure 2.3. Santos Basin chronostratigraphic chart. The Barra Velha Fm (BVE) reservoirs are highlighted by the red rectangle (Modified from Carvalho and Fernandes, 2021 and Moreira et al., 2007)……………………………………………………………………………………..13

Figure 2.4. Crustal domains of the Santos Basin proposed by Zalan (2014) based on Stica (2014). The numbers indicate: Mantle (1), Moho (2), Ductile Lower Crust (3), Conrad discontinuity (4), Brittle Upper Crust (5), Crystalline Basement top (6), Exhumed Mantle (7), and Oceanic Crust (8). Color code for the sedimentary units above the crystalline basement top are: Barremian Syn-Rift (yellow), Early Aptian Syn-Rift (green), Late Aptian Pos-Rift (BVE-light blue), Late Aptian evaporite deposits (purple), Post Ryft Albian Carbonates (dark blue), siliciclastic Late Cretaceous and Tertiary drift sequences (light yellow). The blue rectangle indicates the crustal segment of the Mero Field area, (modified from Zalan, 2014)…………………………………………..…….14

Figure 2.5. Barra Velha Formation main depositional facies at Mero Field (modified from Sartorato et al, 2018). The intrusive igneous bodies are of Santonian age and alkaline signature...14

Figure 2.6. A) Proposed model by Lima et al., (2020) for pre-salt reservoirs of Northern Campos Basin. B) Hydrothermal alteration model proposed by Poros et al., (2017) for Kwanza Basin, Angola…………………………………………………………………………………...….15

Figure 2.7. 3D view of the input seismic cube after cropping around the area of interest…………17

Figure 2.8. 3D view of the base of the Barra Velha reservoir surface (and fault gaps) at Mero Field (left) and Central Libra Appraisal Plan area (right) showing the wells used in this study…………18

Figure 2.9. Workflow used in this study. (RTM = Reverse Time Migration, DST = Drill Stem Test), GLCM = Gray Level Co-occurrence Matrix, K1 and K2 = maximum positive and negative curvatures, HT = Hilbert Transform, SOM = Self Organizing Maps, GTM = Generative Topographic Models)………………………………………………………………………….19

Figure 2.10. 3D view of different seismic attributes used in this study. All of them were extracted within the Barra Velha Formation using its top and bottom horizons. A) Conceptual depositional model for the Barra Velha Formation of the Mero Field proposed by Sartorato et al., (2018). (BU= Build Up, PF= Platform facies, MO= Mounds, EM= Embayment facies, IB= Igneous bodies); B) Depth map of top of Barra Velha surface; C) Hilbert Transform of Amplitude; D) Hilbert Transform

of Fatti´s Lambda-Rho; E) Interval velocity; F) Energy Ratio Similarity; G) GLCM Energy; H) GLCM Entropy; I) GLCM Homogeneity; J) Average frequency; K) Maximum positive curvature; L) Maximum negative curvature……………………………………………………………26

Figure 2.11. A) Hyperdimensional cross plotting: The number of dimensions of clustering is equal to the number of input attributes, in our case from 6 to 8. B) illustration of the basic theory of the clustering algorithms. SOM= Self Organizing Maps, GTM= Generative Topographic Maps. (Modified from Zhao, 2015). ……………………………………………………………….…29

Figure 2.12. 3D view (Inline, Crossline and base of Barra Velha reservoir surface) of four final clustering models for BVE reservoir at the studied area. A) SOM1 using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. B) SOM2 using eight attributes: HTAM, AVF, ERS, GLEN, GLET, GLHO, K1, and K2. C) GTM using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. D) K-means using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. E) 2D color bar of the unsupervised clustering models, use this color bar when examining Figure2.14…………………32

Figure 2.13. Seismic structural map of top of Barra Velha (BVE) reservoir with the studied wells and one lead. The horizontal white lines are the vertical sections paths shown in Figure 2.14……33

Figure 2.14. East-West vertical sections passing through some of the studied wells and one lead. The input amplitude, SOM1, K-means and GTM models are shown. In GTM model vertical sections we interpreted the main features: build ups (BU), mounds (MO), platform facies (PF), igneous bodies (IB), embayment facies (EM) and hydrothermal vents (HV). In the well logs close

to GTM models vertical sections, we have litogeochemical logs (LG) with the highlighted silicon (Si) and calcium (Ca) relative concentrations, and nuclear magnetic resonance (NMR) porosity log. In Well C, the silicon content is related to clay minerals in the platform facies limestones. In Well X4, the silicon content is related both to clay minerals in the embayment facies and silicate minerals in the igneous bodies. In Well X3, the silicon content is related to a pervasive hydrothermal silicification amplifying the metric cave porous system, which is shown in the image log. OWC = oil water contact……………………………………………………………….……………34

Figure 2.15. 3D view of the clustering models using opacity to highlight the main features of interest over the base of Barra Velha reservoir surface and the Oil Water Contact (OWC) plane. A) SOM1 model with opacity to the clusters that are interpreted as sill intrusions. Wells highlighted in yellow penetrated those sills. B) SOM2 model with opacity to the clusters interpreted as sweet spot reservoirs: highly silicified and fractured build ups, mounds and hydrothermal vents. Wells highlighted in yellow penetrated those excellent reservoirs. C) GTM model with opacity to the clusters interpreted as both build ups, mounds, hydrothermal vents and the platform facies. Wells highlighted in yellow penetrated those reservoirs. D)K-means mode with opacity to the clusters interpreted as sweet spot reservoirs: highly silicified and fractured build ups, mounds and hydrothermal vents. Wells highlighted in yellow penetrated those excellent reservoirs………….35

Figure 3.1. A) Location map of the Mero Field and Central Libra Appraisal Plan (CLAP) area; B) Wells and seismic dataset used in this study. Wells A to H have drill stem test (DST) data. Wells

X1 to X5 were drilled for appraisal purposes. Wells A and C were selected as an injector-producer pair to support an anticipated production system test (Modified from Maas et al., 2023)………...49

Figure 3.2. A) Structural section interpreted from original seismic amplitude volume (RTM depth migrated). The section trajectory is plotted over the top of the BVE reservoir surface. Section A-A' passes through wells I (training well, hyper producer), C (training well, good producer), and BT19 (blind test well, highly injective). Well C penetrated highly productive carbonate platform facies (mainly grainstones made by build ups reworking and laminites) close to a hydrothermal vent associated with deep faults. Injection well BT19 was drilled at fault footwall, is highly injective, and the main reservoir is made by gravitational deposits from build up reworking. White arrows indicate paths for hydrocarbons (HC), carbon dioxide (CO2), and silica (Si) rich fluids through reactivated faults that acted as hydrothermal vents. B) Chronostratigraphic chart of Santos Basin: BVE reservoir is composed of sequences K44, K46, and K48 (modified from Moreira et al., 2007). C) Our depositional model of the BVE reservoir at Mero Field (modified from Sartorato et al., 2018)……………………………………………………………………………………52

Figure 3.3. Workflow used in this study. Acronyms: DST (drill stem test), PLT(production logging tool), AASPI (attribute assisted seismic interpretation), SWEET (sweetness), HTAM (Hilbert transform of amplitude), ERS (energy ratio similarity), AVF (average frequency), GLEN (gray level co-occurrence matrix energy), GLET (gray level co-occurrence matrix entropy), GLEH (gray level co-occurrence matrix homogeneity), K1 (maximum positive curvature), K2 (maximum negative curvature, HTLR (Hilbert transform of Fatti's lambda-rho), AVO (amplitude versus offset), FC (flow capacity), PI (productivity index), RF (random forest), KNN (K-nearest

neighbors), SVM (support vector machine), GSCV (grid search cross validation), MLP (multi-layer perceptron)………………..……………………………………………………………55

Figure 3.4. 3D view of the seismic attributes used in this study at the top of Barra Velha reservoir surface. A) Depositional model of BVE reservoir at Mero Field (modified from Sartorato et al., 2018); B) Top of BVE reservoir surface; Seismic attributes: C) Hilbert transform of amplitude; D) Hilbert transform of Fatti's Lambda Rho; E) Sweetness; F) Energy ratio similarity; G) GLCM energy; H) GLCM entropy; I) GLCM homogeneity; J) Average frequency; K) Maximum positive curvature- K1; L) Maximum negative curvature-K2. Geologic features acronyms: BU (Build ups), MO (Mounds), HEP (high energy platform facies), LEP (low energy platform facies), DF (distal facies), HV (hydrothermal vents), (Modified from Maas et al., 2023)……………………………59

Figure 3.5. DST data compilation at BVE formation at Mero Field. This cross plot shows the linear relationship between flow capacity (FC- x axis, logarithmic scale) and productivity index (PI-y axis) obtained through DST. At the left top corner, we have the box plots of FC and PI respectively. Most of the DST data were acquired at mild (8,000 bpd oil production) and good producers (24,000 bpd) from low and high energy platform facies respectively. When closer to hydrothermal vents, these platform reservoirs have higher FC and PI values. The hyper productive wells from silicified build ups and mounds represent anomalies related to very high transmissibility reservoirs characterized by karstified and fractured stromatolitic accumulates. At the Mero field, these sweet spot reservoirs can produce up to 48,000 bpd (BVE only). Note that both FC and PI are not totally

related to depositional facies, the hydrothermal events altered all facies and, as a result, enhanced pre-existing transmissibility, mainly at the build ups and mounds……………………………..…60

Figure 3.6. A) 3D view of the wells used in this study. The surface is from the base of the BVE reservoir with faults. In the Mero Field, the BVE reservoir is deformed by extensional reactivation of rift faults but with relative preservation of the original platform geometry. Wells in red are mostly appraisals and were used for training the supervised machine learning algorithms used in this study. Wells C1 and C2 are pseudo-well locations where BVE has presumed non-reservoir facies. Wells in white are development wells that have both FC and PI data, they were not used for training and, thus, consist of blind tests for quantitative assessment of our prediction models. Wells in pink do not have DST data but have other reservoir quality indicators (mainly mini-DST) and, thus, also can be used as a semi-blind test for qualitative assessment of our predictive models. From B to D, we have 3D views of one example of how the training dataset is built. At the tested interval, we extract seismic mini-cubes (prisms) around the well (nine traces) (Modified from Maas et al.,2023); B) seismic mini-cube of HTAM attribute along the tested interval (145 m) at the BVE reservoir, the sample interval in the z-direction is 5 m. At the perforated zone, we have 30 voxels in the z direction and 9 voxels (3x3 voxels=75x75 m) in the x and y directions, totaling 270 voxels or data points around the perforated interval; C) The same mini-cube has the FC values for each flow unit identified via PLT. D) The mini cube has the values of PI for the same flow units. The surface is from the base of the BVE reservoir. We did the same for the other nine training wells, totaling 4968 training data points……………………………………………….……………62

Figure 3.7. PLT flow units and seismic attributes scale matching. A) Nuclear magnetic resonance (NMR) porosity log at DST interval at Well H; B) Cumulative production curve obtained with production logging tool (PLT) at the same interval; C) Production apportionment obtained with the derivative of the previous curve: the steepest gradients at curve B are related to the main flow units that contribute to the cumulative production. Note that there are some zero flow intervals, where the production at curve B is flat; D) We modulated curve C by both DST values of FC and PI and used them for training dataset building; E) Hilbert transform of input amplitude (HTAM) seismic attribute at well H location: this attribute mimics a seismic bandwidth acoustic impedance. We can observe that the flow units (red polygons) identified through PLT have thicknesses that are compatible with the seismic resolution (40 m); F) Same attribute as in E but with spectral balancing which increased the seismic resolution to 25 m. We also have evaluated the impact of spectrally balanced inputs on our prediction models……………………………….……………………..63

Figure 3.8. A) Heatmap plot of the ten input attributes and flow capacity at the tested intervals of BVE at the training dataset wells. Note that all the input features (attributes) have very low correlation among them. The exception is the GLCM ones, but even though they have correlation values around ± 60% (GLET), which makes them complimentary to each other. Another interesting aspect is the very low correlation between each of the input features with FC, which indicates that with just one attribute at a time, it is impossible to predict FC. B) Seaborn pair plot of some of the normalized input seismic attributes and the flow capacity (hue) at the tested intervals of the training dataset wells. We can observe that most of the input features (except GLCM ones which are relatively highly correlated) are non-linearly separable. It is not possible to separate the hyper flow units (blue dots) from the other ones…………………………………………………………...69

Figure 3.9. Histograms of the ten normalized input seismic attributes and flow capacity values at tested intervals of the training dataset wells. Here we can observe strong overlapping of FC values for each input attribute. Again, with one attribute at a time, we cannot predict FC or PI…………71

Figure 3.10. Some key aspects of the selected ML methods for this study. The illustrations here are simplified to 2D and 3D feature spaces. Keep in mind that in our case everything happens in an 11-dimensioned space. A) Illustration of random forest regression (modified from Goyal, 2022); B) Illustration of KNN regression: Low values of K can lead to overfitting, whereas very high values of K may not capture all the variance. C) Support Vector Machine regression: using support vectors (data points) and an epsilon tube, a hyper plane that best fits the data is built. In the case of non-linearly separable data, the kernel trick is used (D), where a higher dimensional space is emulated, so the hyper plane can be built (modified from Rodriguez Perez and Bajortah, 2022). E) Architecture and main hyperparameters of the MLP used in this study. At the input layer we have 10 seismic attributes at well tested intervals, 5 hidden layers with 600 neurons each (not all of them are shown in the figure), and the output layer with the prediction of flow capacity (FC)………....72

Figure 3.11. FC prediction using random forest (RF) and 10 attributes. A) Geobodies 3D view of the hyper flow units (FC > 100,000 mD.m). Despite not being restricted to depositional facies, these hyper flow units are equivalent mainly to silicified build ups and mounds. At both higher and lower energy platform facies, they are related to hydrothermal vents and tiny mounds. B) Plot of

measured versus predicted FC from random forest training, test $R^2 = 0.84$. C) RF importance of the input attributes. D) SHAP explanation of this predictive model at training dataset…………...75

Figure 3.12. Six of the input attributes at a semi-blind test well location (X1) and RF regression models of FC and PI at BVE Formation. This well penetrated a very high porosity and permeability reservoir related to highly silicified stromatolitic build up. This feature and others can be observed at both input seismic attributes and predicted models of FC and PI. A) Average frequency; B) Maximum positive curvature; C) Maximum negative curvature; D) Energy ratio similarity; E) Hilbert transform of amplitude; F) GLCM entropy; G) Predicted flow capacity; H) Predicted productivity index. I) Composite log of the BVE interval with lithology, NMR porosity, and lithogeochemical spectroscopy log showing silicon concentration (Si). J) Illustration of Well X1 position according to the BVE depositional model (modified from Mizuno et al., 2018). Geologic features acronyms: BU (Build ups), MO (Mounds), HEP (high energy platform facies), LEP (low energy platform facies), IB (igneous bodies), HV (hydrothermal vents)....………………………80

Figure 3.13. RF regression models of FC using six attributes (AVF, K1, K2, ERS, HTAM, GLET) as a transparent overlay on input seismic amplitude (full stack RTM in depth domain). A) EW vertical section passing through wells I (training), BT17 (blind test), and J (training). Close to wells I and BT17, there are image logs showing the cave system which is responsible for the observed very high productivity; OWC= oil water contact; B) Composite log of Well BT17 with lithology, NMR porosity and lithogeochemical spectroscopy; C) EW vertical section passing

through Well G (training); D) Composite log of Well G. This model was predicted using the same attributes as in Figure 13A, but after spectral balancing of original seismic data……………….81

Figure 3.14. 3D view of random forest regression of FC at Mero Field and CLAP area (geobodies) over the base of BVE surface. The hyper flow units related to sweet spot reservoirs (silicified build ups, mounds, and hydrothermal vents) are rendered with an opacity threshold of 100,000 mD.m. At the Central Libra area (CLAP), features like that drilled by Well X5 (sweet spot reservoir) that are above the oil-water contact identified only at the Mero Field (OWC- green plane) could be considered as future targets for oil exploration. Features below OWC can be considered as future injection targets. Hyper flow units away from structural highs are related to either minor mounds or higher permeability, altered platform facies close to hydrothermal vents. There is still a misprediction risk once at 4 out of 20 blind test wells (20%), the FC values were overpredicted (false anomalies)………………………………………………………………………………84

Figure 4.1- Location map of our studied areas in Brazil's offshore Santos. As a training dataset, we used well and seismic data from the Mero field. As blind validation of transfer learning, we used seismic and well data from Bacalhau and Lapa fields (map source: https://geomaps.anp.gov.br). ……………………………………………………………………...101

Figure 4.2 – Schematic geological section showing the general similarities at the Barra Velha (BVE) reservoir between Mero, Lapa, and Bacalhau fields (modified from Zalan, 2024)………101

Figure 4.3 – Illustration of production logging tool (PLT) flow units' distribution according to depositional and diagenetic facies in the Mero field. Wells drilled in the platform exhibit layered flow units related to strongly dolomitized intercalated grainstones, algal mats, and spherulites (low depositional energy). The original depositional porosity is obliterated by superimposed episodes of diagenetic events (dolomitization and silicification). In the build ups and mounds wells (stromatolitic shrubstones) the dolomite cement was almost totally dissolved by silica rich hydrothermal fluids due to very high depositional porosity (interelement), and karst porosity that concentrated corrosive fluids circulation, the overall result is a net permeability gain, sometimes permeability excess (modified from Mimoun and Fernandez-Ibanez, 2023 and Maas et al., 2024)……………………………………………………………………………………102

Figure 4. 4 – a) DST flow capacity, productivity index and production data from the Mero field. The BVE reservoirs can be classified into three major groups: hyper producers silicified build ups and mounds that produce up to 48,000 bpd, good producers form high energy platforms that produce around 24,000 bpd, and mild to low productivity lower energy platform facies producing 8,000 bpd on average (Maas et al, 2024). Second row: scatter plots of (b) flow capacity versus permeability (R^2=0.90) and (c) flow capacity versus daily production (R^2=0.83). Third row: scatter plots of (d) productivity index versus permeability (R^2=0.97) and (e) flow capacity versus daily production (R^2=0.80). Despite being acquired through short periods of either production or injections (transient flows), the flow capacity is an excellent proxy for reservoir permeability prediction at DST and seismic scale, as well as for production forecasting……………………...105

Figure 4.5- General workflow of this research……………………………………………..…109

Figure 4.6- First row: first eigenvector PCA of the nine seismic attributes common to the three different surveys at the BVE reservoir level. The PCA importance of the attributes for eigenvector's variability is persistent across the different surveys. Second row: box plots of the normalized attributes within the BVE zone. The general distribution and skewness are also persistent in the three different seismic surveys. Third row: the correlation heat map pattern is similar across the different seismic surveys. All this PCA and statistical analysis indicate high geological similarity between Mero field (source) and Bacalhau and Lapa fields (target datasets)…………………………………………………………………………………....111

Figure 4.7 – Gross depositional environment (GDE) map of BVE reservoirs at Mero, Bacalhau and Lapa fields. All three areas have the same type of lithologies with different volumetric proportions. At the Mero and Bacalhau fields, the build ups and mounds are expressive. In the Lapa field, the most representative facies are the low depositional energy reservoirs. In the Bacalhau field, there are syn-BVE volcanoclastic mounds that do not occur in other areas. In the Mero field, we have Santonian age sill intrusions in BVE reservoirs that were not yet found in Bacalhau and Lapa fields………………………………………………………..……………113

Figure 4.8- Seismic and well interpretation based geological sections of Lapa field (a), Mero field (b), and Bacalhau field (c). From Lapa to Bacalhau, there is an increasing fault density that affects the BVE reservoir. As a result, the build ups and mounds presence as well as their hydrothermal silicification is more prominent in Mero and Bacalhau fields. This reflects the higher productivity observed in these last two areas………………………………………………………..………..115

Figure 4.9 – 3D view Mero field carbonate platform. The surface is the top of BVE reservoir co-rendered with structural dip to enhance the faults (black shaded areas). Wells in purple are the highest productivity ones (silicified mounds and build ups), wells in blue are mild productive (dolomitized platform facies, locally enhanced near faults), and wells in green are the lowest productivity wells related to dolomitized distal facies. All these ten wells were used as training dataset where their DST flow capacity values were used to train supervised regression algorithms (Maas et al., 2024). At the bottom of the figure, we have the box plots of FC distribution in the Mero high (0 to 350,000 mD.m) and Mero low training subsets (0 to 37,500 mD.m)………..…118

Figure 4.10- Composite log (left), PP synthetic seismogram and HTAM attribute vertical section at one of the Mero training wells location (Well C), which was drilled in the low energy platform. We used DST and PLT flow units from perforations at BVE reservoirs and seismic attributes along with the perforations. As can be seen, the PLT flow units' thicknesses are compatible with seismic resolution, which is 40 m. Because of its zero porosity and proximity to the other perforated zones, we inserted in our training dataset non-measured zero values of FC to counterbalance the biasing of the training dataset which is composed of mostly non-zero values. This can also help our supervised learning algorithms to recognize other types of igneous bodies as nonproductive intervals……………………………………………………………………………………119

Figure 4.11- Top: 3D view of the predictive model of flow capacity (FC) (using random forest regression- RFR) at the Mero field showing some of the training and blind test wells (Maas et al., 2024). The purple geobodies are the predicted hyper flow zones mostly related to heavily silicified

build ups and minor mounds. Middle: same arbitrary seismic section of HTAM attribute with RFR predicted FC. Bottom: same arbitrary section showing SOM clustering unsupervised model and its interpretation. Note the coincidence of high FC zones with the clusters related to build ups and mounds and the low FC zones with clusters related to sill intrusions. For both RFR and SOM clustering, the input seismic attributes used were HTAM, AVF, K1, K2, ERS and GLET……...121

Figure 4.12 – a) Random Forest regression model of FC using all 10 attributes in the Mero high training dataset (with hyper flow units). Both random forest and SHAP attributes' importance show that due to the presence of hyper productive build ups and mounds, the most important attributes are AVF, K1, K2 and ERS due to their convex shape, and frequency attenuation due to intense hydrothermal karstification. The hexagonal bins plot of predicted versus measured FC shows a good fit between the predictive model and the training data (median absolute error = 2.67% and R^2=0.83). b) the same as A but the training dataset is the low productivity Mero subset. In this case both SHAP and RFR attributes importances changed when compared to the Mero high dataset, due to the dominantly layered nature of the reservoir affected by strong dolomitization, leading to lower productivity………………………...……………………………………….122

Figure 4.13 – Mero Field: SOM clustering model using SWEET, HTAM, ERS, K1 and GLEN seismic attributes as inputs. At the blind test well BT8 position, we have the cumulative FC of the tested interval (black rectangle, 125 m thick DST interval, FC=23,500 mD.m). As this well was drilled far away from the major faults, the BVE reservoirs are strongly dolomitized low energy platform facies (higher impedance, red-pinkish clusters), intercalated with less dolomitized layers (lower impedance, green-yellowish clusters). The BT9 well was drilled close to a fault zone, here

the reservoirs are less dolomitized with more cement dissolution due to hydrothermal activity along faults. The FC values in the DST interval (black rectangle, 100 m thick) are the highest ones in the low platform energy (53,000 mD.m). The SOM clustering depicted these hydrothermally enhanced reservoirs as subvertical bluish clusters (fault controlled)…………………………..124

Figure 4.14- Transfer learning from Mero to Bacalhau field: a) Predictive model based on random forest regression (RFR) of flow capacity (FC) at BVE reservoir at the Bacalhau field using local HTAM, AVF, ERS, K1, K2, and GLET seismic attributes inputs. We used the hyperparameters (400 regression trees, max depth=10) of the best RFR model that fitted the Mero field training dataset with hyper productive wells (Mero high) and built a predictive model of FC. At the seismic section passing through B4 and B5 wells from Bacalhau field, we co-rendered HTAM (seismic bandwidth pseudo impedance) and RFR flow capacity prediction. b) Same as A but the overlay is the SOM clustering using the same attributes. c) geobodies from SOM clusters related to the build ups and a vertical section of the RFR of flow capacity model…………………..……………….131

Figure 4.15- Transfer learning from Mero to Lapa field: a) Predictive model based on random forest regression (RFR) of flow capacity (FC) at BVE reservoir at the Lapa field using local HTAM, SWEET, ERS, K1, and GLEN seismic attributes inputs. We used the hyperparameters (500 regression trees, max depth=12) of the best RFR model that fit the Mero field training dataset without hyper productive wells (Mero low) and built a predictive model of FC. At the seismic section passing through all studied wells from Lapa field (a), we co-rendered HTAM and RFR regression model. b) Same as 15a but the overlay is the SOM clustering model using the same seismic attributes. c) 3D view of the highest FC values (>10,000 mD.m) and SOM mounds

geobodies over the middle BVE surface co-rendered with the ERS attribute. Some of the higher FC blobs are associated with faults (dissolution) whereas others are faraway (original permeability remnants)…………………………………………………………………………………...…134

Figure 4.16- Detailed results of RFR of FC and SOM clustering around two of the wells studied in the Lapa field. a) co-rendered HTAM and predicted FC along the perforated interval at Well L4 with the NMR porosity log and PLT flow units. b) co-rendered HTAM and SOM clusters around the same well. c) co-rendered FC prediction and HTAM attributes around Well L2 perforated zones. d) co-rendered HTAM and FC prediction at the same well…………………………...….135

Figure 4.17- Map view of horizon probes of an intra BVE interval (middle of the reservoir) with random forest regression predicted FC values co-rendered with ERS attribute, one of the most important for FC prediction. At the Mero and Bacalhau fields, we can see that the hyper flow units (FC>100,000 mD.m) associated with strongly karstified mounds and build ups are closely associated with lower similarity anomalies (faults, fractures and karst), whereas the lower values are far away from these features. At the Lapa field, the higher FC values are also associated with the faults due to the higher degree of dissolution, but the overall flow capacity is very low……..138

Figure AP1.1 - Crustal domains of the Santos Basin proposed by Zalan (2014) based on Stica (2014). The numbers indicate the: Mantle (1), Moho (2), Ductile Lower Crust (3), Conrad discontinuity (4), Brittle Upper Crust (5), Crystalline Basement top (6), Exhumed Mantle (7), and Oceanic Crust (8). Color code for the sedimentary units above the Crystalline Basement top are:

Barremian Syn-Rift (yellow), Early Aptian Syn- Rift (green), Late Aptian Pos-Rift (light blue), Late Aptian evaporite deposits (purple), Post Ryft Albian Carbonates (dark blue), siliciclastic Late Cretaceous and Tertiary drift sequences (light yellow). The blue rectangle indicates the crustal segment of the Mero Field area, (modified from Zalan, 2014)…………………….…………157

Figure AP1.2 - Santos Basin chronostratigraphic chart. The Barra Velha Fm reservoirs are highlighted by the red rectangle (Modified from Carvalho and Fernandes, 2021 and Moreira et al., 2007)…………………………………..……………………………………………..158

Figure AP1.3 - Early Aptian plate reconstruction of South Atlantic margins showing the extent of the pre-salt Aptian Lacustrine Carbonates and the position of the first two wells which drilled these reservoirs. (Modified from Pietzsch et al., 2018)…………………..…………………………160

Figure AP1.4 - Regional scale well correlation of the Aptian pre-salt carbonates of Santos Basin (Tupi well) and Campos Basin (Whales Park well), see Figure. 3. (Modified from Carminatti et al., 2018)…………………………………………………………...……………………161

Figure AP1.5- Third order sequence proposal for Campos Basin Late Aptian carbonates (Dias et al., 2005). A) Dendritic bush shaped stromatolites; B) Dendritic stromatolites; C) Bush shaped stromatolites. This classification used the terminology proposed by Grotzinger and Knoll (1999): D) Outcrop photograph of dendritically branching tufa (light) which forms stratiform sheets on the scale of1–2 cm. Branching structures are infilled with marine cement and micritic sediment. E) Photomicrograph of a single dendrite structure with thin central stalk and numerous broad branches, both formed of micrite (dark). Lighter material is void-filling marine cement. Note arrow indicating growth orientation. (Modified from Dias et al., 2005 and Grotzinger and Knoll, 1999)…………………………………………………………………………………………161

Figure AP1.6 - Depositional model for Barra Velha Fm carbonates at Santos Basin. The deposition of the "microbialites" was strongly controlled by the substrate paleo relief…………………….….163

Figure AP1.7- Late Aptian plate reconstruction showing that the Barra Velha carbonates were deposited in a Continental paleo lake (Tupi Lake) which was isolated from Proto South Atlantic Ocean (SSA) by the Walvis Ridge. No marine fossils were found within the Tupi paleo lake. (Modified from Arai, 2014)………………………………..…………………………………165

Figure AP1.8- Abiotic induced precipitation model proposed by Wright and Barnett (2015)…...166

Figure AP1.9- Depositional model proposed by Souza et al., (2018). The deposition of BVE carbonates was dominantly controlled by abiotic factors……..………………………………..166

Figure AP1.10- CO2 concentrations in the Santos Basin pre-salt accumulations (Santos Neto et al., 2012)…………………………………………………………………………..………..168

Figure AP1.11- Hydrothermal silicification al the Pao de Açucar Field at Campos Basin. A) Geological section showing the main reservoir units; B) Thin section photo of silicified deep lake spherulitites; C) Silicified hydraulic vein; D) Late dolomitization of a hydraulic breccia; E)

Heavily silicified microbial carbonate with opal vugs; in all thin section images the scale bar is 200 micrometers. (Modified from Vieira de Luca et al., 2017)……………………...…………..169

Figure AP1.12- Pore size distribution after hydrothermal silicification at the Pao de Açucar carbonates. The pore throat sizes were augmented at the (A) micro, (B) meso and (C) Giga-pore (crystal caves) scales at image log. (Modified from Vieira de Luca et al., 2017)……………..….170

Figure AP1.13- A) Hydrothermal silicification model and B) Diagenetic history proposed by Poros et al., (2017) for the Kwanza basin Aptian lacustrine carbonates………………………………..171

Figure AP1.14- Hydrothermal events sequence at northern Campos Basin lacustrine carbonate reservoirs. (A) Syngenetic crust of coalesced fascicular calcite aggregates covering syngenetic Mg-clay matrix. B) Microcrystalline and blocky dolomite partially replacing the Mg-clay matrix, calcite spherulites, and fascicular aggregates. C) Microcrystalline quartz and chalcedony spherulites partially to fully replacing preexisting constituents, and filling primary and secondary porosity. D) Silicification of some dolomitized fascicular and spherulitic aggregates. E) Intense silicification, fracturing and dissolution. Saddle dolomite partially filling fracture and vugular pores. F) Macrocrystalline calcite, radiated Sr-barite and celestine, fibro-radiated zeolite, euhedral to subhedral pyrite, chalcopyrite, galena and sphalerite, and massive bitumen filling the secondary pores. The numbering is in accordance with the paragenetic sequence in Figure. AP1.15………173

Figure AP1.15- Sequence of diagenetic and hydrothermal events affecting the lacustrine carbonate reservoirs at northern Campos Basin. (Lima et al., 2020)…………………………….………174

Figure AP1.16- Proposed model (Lima et al.,2020) for the origin of the hydrothermal alteration of the Aptian lacustrine carbonate reservoirs at northern Campos Basin. The fault-focused hydrothermal system probably involved mixing of fluids derived from several sources: (B) Pre-Cambrian basement (mainly granitic-gneissic felsic rocks), (C) volcanic rocks from the Cabiúnas Formation (mainly basaltic), (R) rift deposits from the Atafona and Coqueiros formations, and (S) serpentinization of the upper mantle. Note the intrusive magmatic (mafic) rocks (M) in the Coqueiros (rift interval) and Macabu (sag section) formations; hydrothermal alteration (H) comprising extensive dolomitization, silicification, and dissolution, with the paragenesis including saddle dolomite, macrocrystalline calcite, mega-quartz, Sr-barite, celestine, fluorite, dickite, sphalerite, galena, and other metallic sulfides filling fractures and dissolution porosity; and anhydrite thick layer (A) at the top of the sag reservoir affected by the hydrothermal alteration (base of evaporites)……………………………………………………………………………..175

Figure AP1.17- Petrographic classification scheme proposed by Gomes et al., (2021) for the Barra Velha lacustrine carbonates…………………………………………………………………..177

Figure AP1.18- Depositional model proposed for the BVE carbonates at Mero Field based on well and seismic data (modified from Sartorato et al., 2018)…………..……………………………..178

Figure AP1.19- Travertine towers at the Lake Abhe, Djibouti (Awaleh et al., 2015)………….....180

Figure AP1.20 – Outcrop study of Eocene carbonates from Green River Fm (Cupertino, 2023). Bioherm facies succession: A) Individuals bioherms, west of Rosewood Section, showing distinctive domical shape. B) distinctive facies succession within bioherms: fine-grained domical stromatolite facies (M3), followed by arborescent (M4) and branched minicolumnar stromatolites (M5), columnar stromatolites (M6), and Chlorellopsis (M7) facies at the top. Non-microbialite facies: Intraclastic conglomerate facies (NM1) and Ostracodal grainstone facies (NM4)………181

Figure AP1.21- The link between tectonic evolution and fluid migration along the Alpine Tethyan margin. The Grosina and Bernina detachments record a fluid with a continental composition (about 190-185 Ma). Between 185 and 165 Ma, the ED fault rocks and the supradetachment basin record mantle-related fluids that are produced by initial serpentinization under thinned continental crust. Serpentinization and fluid flow evolve with the progressive mantle exhumation toward the seafloor. (Extracted from Pinto,2014)…………………………………………….…………….183

Figure AP1.22 - Numerical modeling results of Aptian evolution of Santos-Kwanza conjugate margins. (Araujo et al., 2022)……………………………………………….……………………..184

Figure AP2.1 – Random Forest composed of multiples decision trees; the maximum depth is the one composed of only leaf nodes. The final prediction is done by averaging the leaf nodes of each decision tree. Very shallow trees tend to underfit whereas too deep trees tend to overfit the data (modified from Goyal, 2022)…………………………………………………………………193

Figure AP2.2- Max depth =1 decision tree (light blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y. In this kind of scatterplot, the data density (color coded, white for low density and purple for high density) is shown within each one of the regularly spaced hexagonal bins over the 5000 datapoints. It is an interesting way to see regression results with such a high number of datapoints…………….. …194

Figure AP2.3- Max depth =2 decision tree (mustard line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…....195

Figure AP2.4- Max depth =3 decision tree (mustard line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…....195

Figure AP2.5- Max depth =5 decision tree (red line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……….196

Figure AP2.6- Max depth =10 decision tree (orange line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y….196

Figure AP2.7- K =10 KNN regression model (blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……197

Figure AP2.8- K =50 KNN regression model (blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…....198

Figure AP2.9- K =100 KNN regression model (purple line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…....198

Figure AP2.10- K =1000 KNN regression model (red line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…....199

Figure AP2.11- K =2000 KNN regression model (orange line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y………………………………………………………………………………………....199

Figure AP2.12- Illustration of the kernel trick which was applied to linearize the data and then apply the SVR using the support vectors (ξ, blue arrows) and epsilon parameters (+Ɛ, - Ɛ), after Bock et al., (2020)……………………………………………………..………………..200

Figure AP2.13- SVR using epsilon= 0.1 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…………………….201

Figure AP2.14- SVR using epsilon= 0.5 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…………………….202

Figure AP2.15- SVR using epsilon= 0.8 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…………………….……202

Figure AP2.16- SVR using epsilon= 1 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…………………………..203

Figure AP2.17- MLP neural network architecture that best fits the data of this toy example. It is composed of one input layer of x values (5000 datapoints), 3 hidden layers with 10 neurons each, and an output layer with the prediction value that are closest possible to the analytic solution (y=sin(x)). For real world problems (like the ones described in this dissertation), more complex architecture is used such as: input layer with multiple features (or seismic attributes), up to dozens of hidden layers made of hundreds of neurons each. In the case of regression, the output layer is the single numerical prediction for each input layer datapoint…………………………………..204

Figure AP2.18- MLP regression using one hidden layer with ten neurons to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……………………………………...……………………………..205

Figure AP2.19- MLP regression using two hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y…………………………………………………………...………..205

Figure AP2.20- MLP regression using three hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……………………………………………………………………….....206

Figure AP2.21- MLP regression using five hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……………………………………………………………………...…..206

Figure AP2.22- MLP regression using ten hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y……………………………………………………………………….207

List of Tables

Table 2.1 – Key metrics of the Mero Field (after Carlotto and Silva, 2018)………………………15

Table 2.2 – A) Seismic acquisition parameters; and B) Geometry of the cropped version of the velocity and reverse time migrated depth volumes……………………………………...…..17

Table 2.3 – Reservoir quality data compilation for wells (wildcat and appraisal) drilled at Mero Field and Central Libra appraisal plan area for the Barra Velha Formation (BVE). Wells in blue are those drilled in the sweet spot reservoirs: silicified shrubby stromatolites within build up facies…………………………………………………………………………………..22

Table 2.4 – Correlation matrix of selected attributes within the Barra Velha Formation at the studied area. The closer to 1 or -1, the higher the correlation. The values shown in red are the high correlation among the redundant GLCM´s attributes……………………………….…………24

Table 2.5 – Attributes signatures related to each facies of the Barra Velha Formation at the studied area……………………..………………………………………………………………………25

Table 3.1–A view of the Pandas data frame of the training dataset. Here we have all data points information of normalized seismic attributes and flow capacity at the BVE reservoir tested intervals of the Mero Field………………………………………………………………..64

Table 3.2- Summary of the results of supervised machine learning (ML)regression of flow capacity (FC) and productivity index (PI) for BVE reservoirs at Mero Field and Central Libra block.

Acronyms: GSCV (grid search cross validation), SB (Spectral Balancing); RBF (radial basis function), MLP (multi-layer perceptron), ReLU (rectified linear unit)………………………...……76

Table 4.1- Main parameters of the three different seismic surveys used in this study. Due to technical limitations of early 2000's marine seismic acquisition, seismic data over the Bacalhau field area was acquired with shorter cable length (6200 m) when compared to Lapa and Mero fields (>8000 m), for this reason, in the Bacalhau field, amplitude versus offset (AVO) data is compromised and, so, we did not use AVO attributes in this study. Acronyms: RTM (reverse time migration), KPSDM (Kirchhoff pre-stack depth migration)…………………………………...…..104

Table 4.2- Results summary of random forest regression of fault capacity transfer learning from Mero to Bacalhau and Lapa fields. The two first rows are the models fitted to training subsets with (Mero high FC - Bacalhau like) and without (Mero low FC- Lapa like) hyper productive wells from Mero Field. The third and fourth rows are respectively the predictive models in the Mero high FC and Mero low FC (red cube in Figure 4.12). The fifth and sixth rows are the predictive models using transfer learning to Bacalhau and Lapa fields. The overall blind test performance in the three seismic surveys is around 80%, which is very close to the GSCV performances during training. This is a very good indication of the high generalization capacity of our predictive models trained with Mero field data……………………………………………………..………………127

Table AP1.1 - Carbonate rock classification scheme proposed by Terra et al., (2010)…………..164

Abstract

Even almost 20 years after the discovery of the prolific pre-salt oil province in SE Brazil offshore, seismic characterization of these complex reservoirs is still a challenging task. In the oil and gas fields, where dozens of wells are available, the seismic inversion-based workflows are proven to be the best for reservoir characterization. However, in exploratory stage areas the reduced number of wells makes seismic inversion non-viable. As a result, solely seismic amplitude-based workflows are strongly affected by ambiguity pitfalls, which delay the appraisal phase, postpone commercial production, and, thus, erode the economic value of the project. As a solution, I propose a machine learning approach that leverages the big amount of dynamic reservoir data in a known field (Mero field in Santos Basin) and easy-to-obtain post-sack 3D seismic attributes, which are common to both production and exploration areas and are available as soon as the seismic processing is done, without requiring wells for calibration like seismic inversion does. Well dynamic data, mainly drill stem tests, are the most reliable information on reservoir productivity before the commercial production and are available during the upstream phase. In this sense, my doctoral research was structured in three main projects: the first one (Chapter 2) on qualitative reservoir modelling using post-stack seismic attributes and unsupervised learning techniques. I used combinations of ten different seismic attributes (interval velocity, amplitude versus offset, complex traces, geometric and voxel-based texture) as inputs for self-organizing maps, generative topographic mapping and k-means clustering algorithms for seismic facies discrimination in the Barra Velha reservoir of the Mero field. The seismic facies models were validated using information from 13 wells which were not used for any training. The best models were obtained with self-organizing maps.

The second project (Chapter 3) is on quantitative interpretation using the same attributes and drill stem test data to train supervised learning algorithms for prediction of reservoir productivity of the Barra Velha carbonates in the Mero field. I tested classic supervised learning algorithms (shallow learning) of random forest, support vector machines and K-nearest neighbors for supervised regression of flow capacity and productivity index. I also tested a deep learning algorithm (multi-layer perceptron) to compare its cost-effectiveness to the shallow learning algorithms. For the validation of my predictive models, I used information from 20 blind test wells. The best results were obtained with random forest regression of flow capacity (85% blind test performance).

In the third project, I studied how transfer learning techniques can leverage the machine learning training using Mero field data to accurately predict reservoir facies and productivity in other seismic surveys 200 km distant from Mero field: the Bacalhau and Lapa fields, which have production data to validate my predictive models. Using self-organizing maps and random forest algorithms trained with Mero field data, I could accurately predict (80% average performance) the facies distribution and flow capacity values observed in the blind test wells in Bacalhau and Lapa fields. Transfer learning using Mero training proved effective for reservoir de-risking and upstream optimization even when working with multiple seismic surveys.

As I used post-stack seismic attributes, well test productivity, and injectivity data from subsurface reservoirs to train our models, this approach can be used in any kind of project such as CCUS, geothermal, and hydrogen storage projects in the context of the energy transition.

Chapter 1: Introduction

In deep water settings, where all operations are more complex and expensive, the state of art of reservoir characterization assisted with 3D seismic data is the application of seismic inversion for accurate facies distribution and reservoir property modelling (Mitra and Khara, 2001; Taboada et al., 2001; Raeesi et al., 2012; Alfarraj and Alregib, 2018; Abdulaziz et al., 2019; Anifowose et al., 2019; Hosseini et al., 2019, Penna and Lupinacci,2020). However, accurate seismic inversion cubes require dozens of wells for a more robust wavelet estimation and low frequency information. Moreover, in deep water settings, this can take more than ten years to be available and will be useful only for field development, whereas during exploration phases there is limited or even no well information. The most common type of seismic data available are post stack seismic attributes and, in some cases, amplitude versus offset (AVO) data. In these operationally challenging environments, increased time spent between field discovery and commercial oil production, decreases the economic value of the project. These drawbacks can be overcome by leveraging the fact that the petroleum industry has a huge amount of data from a plethora of different production assets (big data), that could be used for reservoir characterization during the upstream phase in virgin areas (greenfield exploration).

In this sense, I propose that we can use production information (well test data and daily production rates) of well-known assets, post stack seismic attributes, and machine learning (ML) techniques to bypass the seismic inversion process and make accurate predictive models of facies distribution and reservoir productivity. This is all to optimize and accelerate the upstream phase by reducing risk and, thus, adding economic value. The main work hypothesis is that the observed seismic trace is a result of the interactions of the wavefront and different kinds of lithologies and even different permeability-porosity domains within a reservoir. As a result, seismic attributes enhance these reservoir heterogeneities that are not obvious in the amplitude volume. Furthermore, these heterogeneities must be observable at the seismic scale (bigger than seismic resolution). In this sense, seismic attributes and well dynamic data should be suitable for machine learning assisted reservoir characterization (lithology and productivity modeling). The main research question that motivates my PhD research is: **"Do post-stack seismic attributes significantly predict reservoir facies distribution and well productivity during upstream phases?"**

To answer this broad research question, this dissertation is structured as follows:

Chapter 2 – The secondary research question that motivates this chapter is: **"Can unsupervised clustering analysis of post-stack seismic attributes effectively characterize reservoir facies distribution and productivity?".**

To address this issue, I used seismic attributes (interval velocity, instantaneous, geometric, texture, and AVO) and three different unsupervised clustering methods (self-organizing maps-SOM, generative topographic mapping-GTM, and K-means) to build a predictive facies and productivity model of a well-known carbonate reservoir from a producing oilfield in Brazil offshore deepwater (Mero field). I used information from existing wells that were not used for any training to validate the unsupervised clustering models.

Chapter 3 - The secondary research questions that motivates this chapter are: **"Can a supervised learning model using multiple seismic attributes and well dynamic data accurately predict inter-well reservoir productivity, and does deep learning outperform traditional supervised methods in this prediction task?".**

To answer these questions, I used drill stem test (DST) data from ten wells of the Mero field and post stack seismic attributes (instantaneous, geometric, and voxel-based texture) to train supervised regression algorithms for prediction of carbonate reservoir flow capacity (FC) and productivity index (PI), two important DST parameters. I focused on the most classic supervised regression methods: random forests (RF), k-nearest neighbors (KNN), and support vector machines (SVM). Additionally, I tested a deep learning algorithm (multilayer perceptron – MLP) to investigate its cost effectiveness when compared to the classic methods. My FC and PI predictive models were validated using information from 20 wells that were not used for any training (blind test validation).

Chapter 4 - The secondary research question that motivates this chapter is: **"Can machine learning models trained on Mero field data be effectively transferred to areas with limited well control, and does this transfer learning approach demonstrably reduce risk and optimize upstream project outcomes?**

From previous research described in chapter 3, I chose the best ML algorithms trained and validated with Mero field data and tested them in two other oil fields of the Santos basin hundreds of kilometers distant from Mero and covered by different seismic surveys acquired and processed in different times and using different technologies. The transfer learning in this case consists of using the same seismic attributes combination and same SOM and RFR hyperparameters of Mero field to make predictions in other areas distant from the Mero Field and with roughly similar geology (same depositional system, same geological age, same basin). I show how I used SOM clustering and RFR trained with Mero field data and transferred it to make accurate predictions in the Lapa and Bacalhau oil fields that also have blind test wells that were used for my transfer learning models validation. The main challenges are the different seismic surveys and local geology discrepancies such as non-productive volcanoclastic mounds in the Bacalhau field that are not present in the training dataset.

Chapter 5 – I summarize the main conclusions based on the results described in the Chapters 2, 3, and 4 of this dissertation.

Appendix 1 – It is a study of the state of art of the geological knowledge evolution on the main reservoir unit described in this dissertation: the unique Aptian lacustrine carbonates of the Barra Velha Formation (BVE) of southeastern Brazil offshore (Pre-Salt). I describe how the geological understanding of the BVE reservoir has evolved from the first wells drilled in the early 2000' to the present day. It is a good opportunity for the reader to get up to date about the main geological aspects of these unique high yield reservoirs.

Appendix 2 – I present toy examples of supervised regression of one of the simplest nonlinear functions (y = sin (x) + 20% noise). I show how the main hyperparameters of decision trees (the building blocks of random forests), k-nearest neighbors, support vector machines, and multilayer perceptron affect the models' mean absolute error and R^2. The main purpose of this appendix is to show to the reader how these algorithms work, for the sake of simplicity, on low dimensional, nonlinear separable data. The way they work with real world problems is similar, as data are mostly nonlinear separable but with high dimensionality (multiple features as inputs).

References

Abdulaziz, M. A., Mahdi, H. A, Sayyouh, M.H. 2019. Prediction of reservoir quality using well logs and seismic attributes analysis with an artificial neural network: A case study from Farrud

Reservoir, Al-Ghani Field, Libya, Journal of Applied Geophysics, Volume 161, Pages 239-254, https://doi.org/10.1016/j.jappgeo.2018.09.013.

Alfarraj, M., Alregib, G. 2018. Petrophysical-property estimation from seismic data using recurrent neural networks, SEG Technical Program Expanded Abstracts: 2141-2146. https://doi.org/10.1190.

Anifowose, F., Abdulraheem A., Al-Shuhail, A. 2019. A parametric study of machine learning techniques in petroleum reservoir permeability prediction by integrating seismic attributes and wireline data, Journal of Petroleum Science and Engineering, Volume 176, Pages 762-774, https://doi.org/10.1016/j.petrol.2019.01.110.

Hosseini, M., Riahi, M. A., & Mohebian, R. 2019. A Meta attribute for reservoir permeability classification using well logs and 3D seismic data with probabilistic neural network. Bollettino di Geofisica Teorica ed Applicata. 60. http://dx.doi.org/10.4430/bgta0257.

Mitra, P. P., Kharak S. 2001. Synergy through Seismic Attributes in Reservoir Management-a Case Study. Paper presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia, doi: https://doi.org/10.2118/68648-MS.

Penna, R., Lupinacci, W. 2020. Decameter-Scale Flow-Unit Classification in Brazilian Presalt Carbonates. SPE Reservoir Evaluation & Engineering. 23. 1-20. 10.2118/201235-PA.

Raeesi, M., Moradzadeh, A., Ardejani, F.D., Rahimi, M. 2012. Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks, Journal of Petroleum Science and Engineering, Volumes 82–83, Pages 151-165, https://doi.org/10.1016/j.petrol.2012.01.012.

Taboada, R., Condat, P., Corsini, V., Mir, E., Conti, J., Fortunato, G., Villivar, O., List, D., Stancel, S., and L. Jones. 2001. El Tordillo Reservoir Static Characterization Study: El Tordillo Field, Argentina. Paper presented at the SPE Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, March 2001. doi: https://doi.org/10.2118/69660-MS.

Chapter 2: Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data*

*This study is published in the Marine and Petroleum Geology journal as

Maas, M. V. R., Bedle, H., Matos, M. C. 2023.Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data. Marine and Petroleum Geology, Volume 151, https://doi.org/10.1016/j.marpetgeo.2023.106199.

Abstract

The pre-salt high-yield carbonate reservoirs of Brazil are responsible for about 75% of national oil production. Still, seventeen years since its discovery, detailed seismic characterization of these reservoirs remains challenging, with producing and nonproducing lithologies often exhibiting the same amplitude response. Those reservoirs are complex, composed of high productivity lacustrine carbonates (stromatolites and grainstones), mixed with non-producing mudstones, clays, and igneous intrusions. To address this heterogeneity, we evaluate unsupervised machine learning clustering techniques including Self Organizing Maps (SOM), Generative Topographic Maps (GTM), and K-means and apply them to a suite of instantaneous, geometric, interval velocity, and AVO seismic attributes. We use both static and dynamic reservoir data from 13 wells at the Mero field that penetrated lacustrine carbonates from Barra Velha Formation to validate the accuracy of our unsupervised learning models. These wells are excellent blind tests as they were not used for any algorithm training. In this pioneer work, the first done using machine learning in the Mero field, we find that SOM, GTM, and K-means models provide clusters that can be linked to sweet spots associated with fracturing, karstification, and hydrothermal alteration. Additionally, clusters related to sill intrusions, platform facies and distal fine grained non-

reservoirs were identified. The results can be extrapolated to the adjacent Central Libra Area (still in exploration phase), where we have identified the same clusters of the sweet spots reservoirs in Mero Field and, thus, they can be considered as future targets. For accuracy assessment, the same attribute combination SOM clustering was performed in another of Petrobras' pre-salt confidential dataset and the results were satisfactory. So, for exploration optimization purposes, the new proposed workflows can also be applied to other areas with geological similarity to the Mero Field.

Introduction

Seventeen years after its discovery, seismic prediction of Brazilian pre-salt carbonate reservoirs remains a difficult task, with the primary challenge related to the ambiguous response between producing and non-producing lithologies. For instance, one interpretation pitfall is the very similar amplitude signature observed from permeable grainstones and laminites of the platform facies and in non-reservoir mudstones of the embayment facies when both are under superimposed tectonic deformation. Furthermore, high productivity facies do not always have a clear amplitude response due to seismic imaging issues. This can be observed in the excellent reservoirs (shrubby stromatolites) of the carbonate build ups and mounds, which are poorly illuminated beneath overlying complex salt bodies. These limitations often postpone commercial production if they lead to suboptimal well location during the appraisal and development phases.

In deep water E&P projects, the increased time spent, and extra wells drilled during the appraisal and development phase decreases the final economic value of the project. In this context, machine learning (ML) assisted seismic interpretation promises to speed up all the phases by helping to more rapidly identify not only the reservoir sweet spots, but also other features of interest, like igneous bodies which can give rise to both drilling and flow problems. When seismic multi-attributes are derived and analyzed together, these features become more easily detectable (Chopra and Marfurt, 2007). However, for human interpreters, it is very time consuming to analyze each one of these attributes separately for facies classification. Unsupervised clustering techniques, in turn, can analyze multi-attributes in a high dimensional space and, thus, identify different clusters that have similar characteristics which can be extrapolated to carbonate and other lithologies. Unsupervised clustering also accelerates exploratory data analysis, highlighting seismic facies even in areas affected by seismic noise. Although attribute selection has strong human interaction and thus bias, these techniques provide a relatively unbiased view of the subsurface.

This paper describes the first work on machine learning in the Mero field and adjacent areas. We explored Self Organizing Maps (SOM) (Kohonen, 1982 and Matos et al., 2010), Generative Topographic Maps (GTM) (Bishop, 1998 and Zhao et al., 2015), and K-Means (Forgy,1965; Lloyd,1982; and Zhao et al., 2015). Seismic facies characterization with similar methodology has been attempted in deep water turbidites (La Marca & Bedle, 2022). We also enhance the analysis by comparing SOM, GTM and K-means results as well as the variations among these methods when different sets of input attributes are employed.

Area Location and Geological Background

The studied area is the Mero Field in Santos Basin deep waters (Southeastern Brazil) (Figure 2.1). It is a world class asset that was discovered in 2010 by Petrobras's wildcat drilling for Brazilian Petroleum Agency. In 2013, a joint venture composed of Petrobras (operator), Shell, TotalEnergies, CNPC, CNOOC, and PPSA companies carried out the operation. Since then, thirteen appraisal and twenty-five development wells have been drilled and commercial production started in April of 2022. The Mero field is adjacent to the Central Libra Appraisal Plan, where only two exploratory wells have been drilled to date. This configuration is interesting because any pattern detected by machine learning at Mero Field can be immediately extrapolated to Central Libra once they are covered by the same seismic dataset (Figure 2.2).

The Santos Basin is a passive margin basin that evolved from a continental rift that began in the Early Cretaceous (Neocomian/Barremian) (Figures 2.3 and 2.4). The basin started after mantle plume-driven lithospheric stretching with the rift phase half grabens concentrated in the stretched and thinned continental crust domains during the Barremian until the Early Aptian (Zalan, 2014., Stica, 2014) (Figure 2.4). In these grabens, there are, from bottom to top: basic volcanic rocks, talc-stevensite sequences, lacustrine source rocks, and coquina deposits. In the Late Aptian, after the rift phase dominated by mechanical subsidence, a sag basin was developed with the deposition of lacustrine carbonate reservoirs of the Barra Velha Formation (BVE – Figure2.3) mainly in flat platforms over basement horsts and build ups and mounds related to deep faults. At the end of the Late Aptian, when the oceanic stage began, a thick evaporite sequence was deposited, followed by a typical transgressive marine drift sequence during the Late Cretaceous and Tertiary. During the

Santonian, an alkaline magmatic event is responsible for the intrusion of mainly sills within BVE reservoirs and overlying evaporites (Moreira et al, 2007 and Carvalho and Fernandes, 2021).

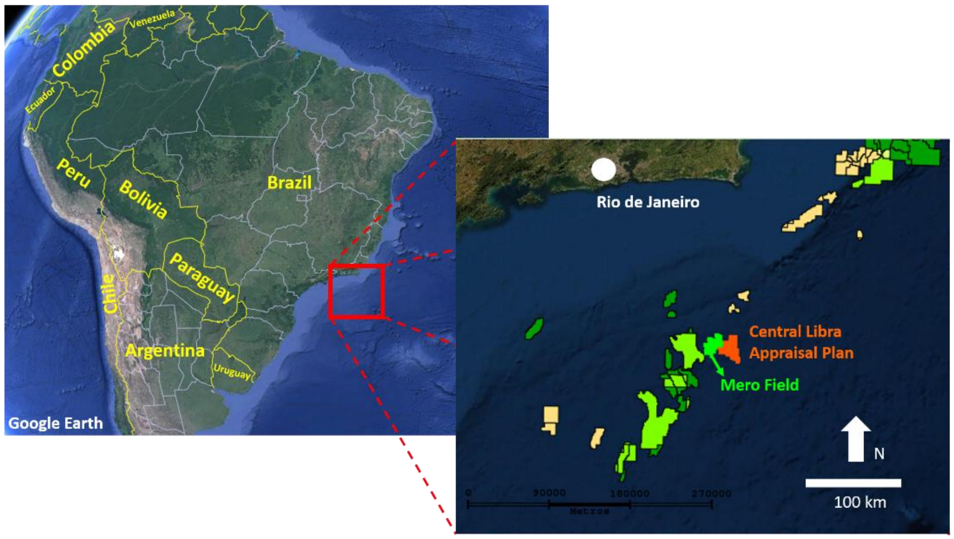

Figure 2.1. Location map of the Mero Field and Central Libra Appraisal Plan.

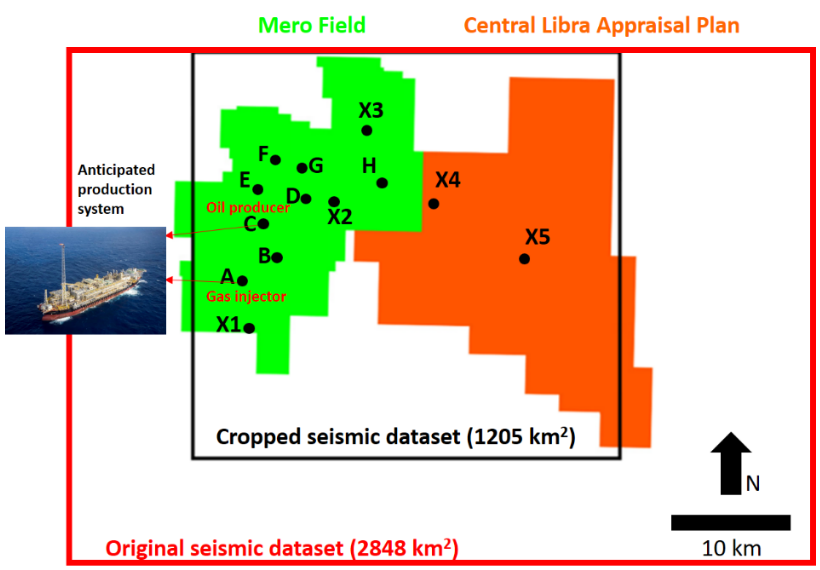

Figure 2.2. Location map of the Mero Field and Central Libra Appraisal Plan with the wells and seismic dataset used in this study. Wells A to H have drill stem test (DST) data. Wells X1 to X5 were drilled for appraisal purposes. Wells A and C were selected as an injector-producer pair to support an anticipated production system test.

In the Mero field, the main reservoirs are Late Aptian post rift lacustrine carbonates of BVE (Figure 2.3), that is the main object of this paper. Secondarily, syn-rift coquinas from Itapema Formation (Figure 2.3) are also hydrocarbon producers (Carlotto and Silva, 2018). Within BVE, the main producing facies are carbonate build ups (BU) and mounds (MO) made by shrubby stromatolites and minor travertines related to deep faults over basement horsts. Other important producing lithologies are their associated platform facies (PF) where they are reworked as grainstones and rudstones intercalated with algal laminites and spherulitites. In the structural lows, there are embayment facies (EM) composed mainly of distal mudstones and magnesium rich clays which are non-reservoirs. All these units are commonly intruded by Santonian alkaline igneous bodies (IB) (mostly sills and minor dykes) (Sartorato et al., 2018) (Figure 2.5). Table 2.1 summarizes the key elements of the Mero Field reservoirs.

Another peculiarity of this area is the pervasive hydrothermal alteration of the reservoir characterized by moderate to strong silicification mainly within the build ups and mounds close to deep faults. Lima et al., (2020) showed petrographic and geochemical evidence of hydrothermal alteration of pre-salt lacustrine carbonate reservoirs in the Northern Campos Basin (Figure 2.6A). Tritla et al. (2013) identified the same kind of hydrothermal alteration within these carbonates in the Southern Campos Basin. Poros et al., (2017) identified the same processes at the conjugate margin in Kwanza Basin, offshore Angola (Figure 2.6B). It seems that these hydrothermal processes occured at a regional scale at both South Atlantic conjugate margins. It will be shown in the subsequent sections that, at Mero Field, this hydrothermal alteration has a strong positive effect on the reservoir productivity. The most productive wells are located in strongly silicified and karstified carbonate build ups and mounds. It will also be shown that hydrothermal vents, silicified build ups and mound-like features are often seen in some seismic attributes and unsupervised clustering models.

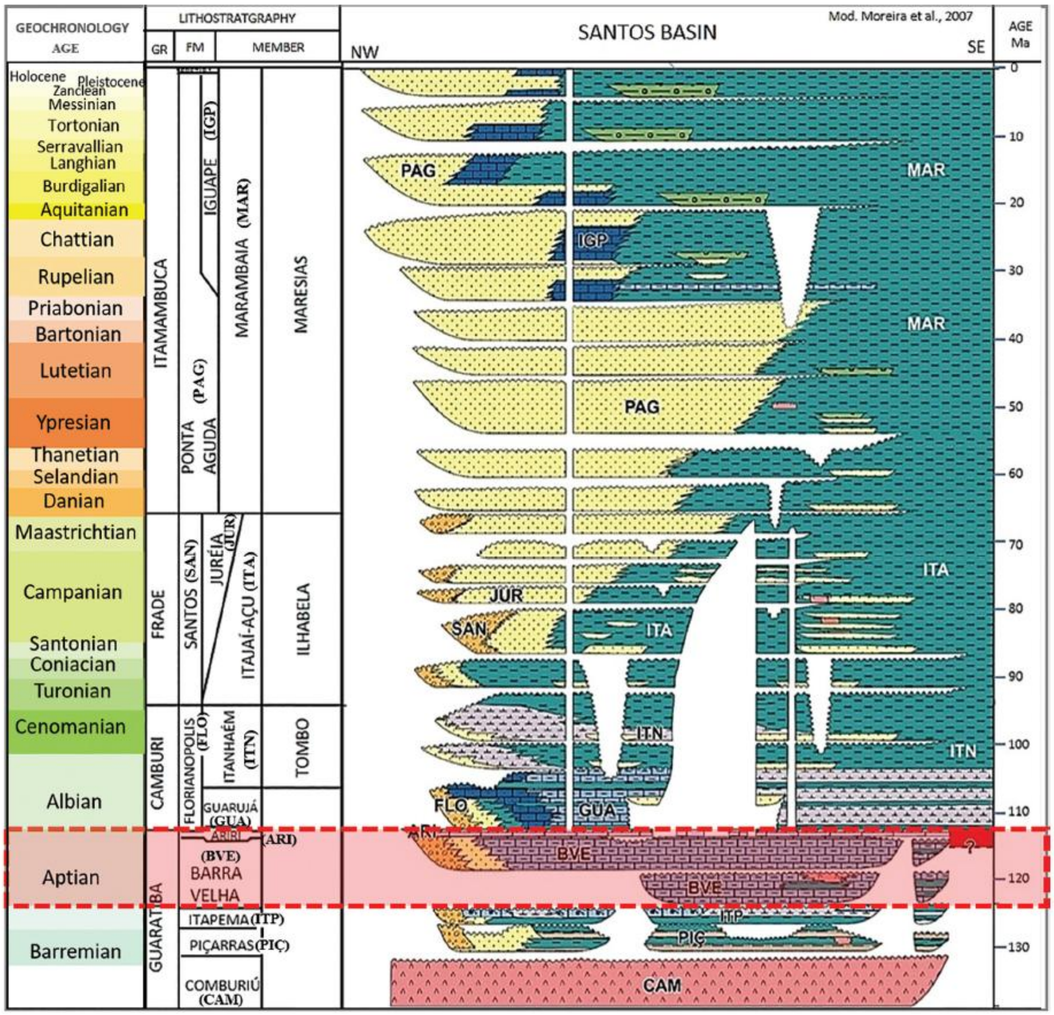

Figure 2.3. Santos Basin chronostratigraphic chart. The Barra Velha Fm (BVE) reservoirs are highlighted by the red rectangle (Modified from Carvalho and Fernandes, 2021 and Moreira et al., 2007).

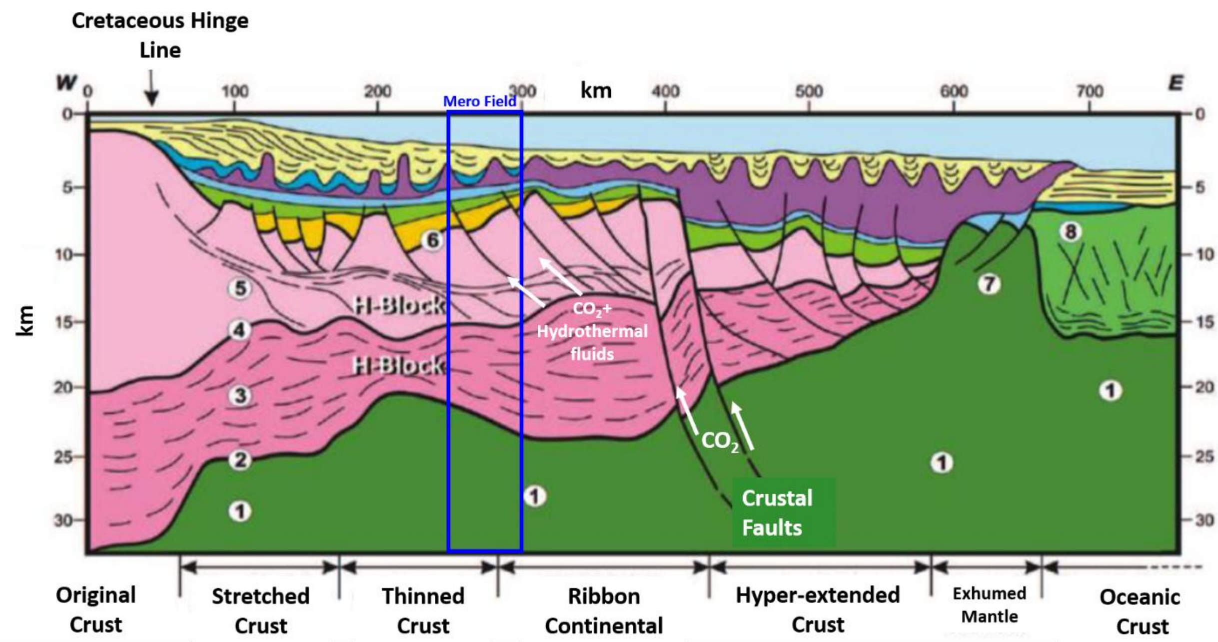

Figure 2.4. Crustal domains of the Santos Basin proposed by Zalan (2014) based on Stica (2014). The numbers indicate: Mantle (1), Moho (2), Ductile Lower Crust (3), Conrad discontinuity (4), Brittle Upper Crust (5), Crystalline Basement top (6), Exhumed Mantle (7), and Oceanic Crust (8). Color code for the sedimentary units above the crystalline basement top are: Barremian Syn-Rift (yellow), Early Aptian Syn-Rift (green), Late Aptian Pos-Rift (BVE-light blue), Late Aptian evaporite deposits (purple), Post Ryft Albian Carbonates (dark blue), siliciclastic Late Cretaceous and Tertiary drift sequences (light yellow). The blue rectangle indicates the crustal segment of the Mero Field area, (modified from Zalan, 2014).

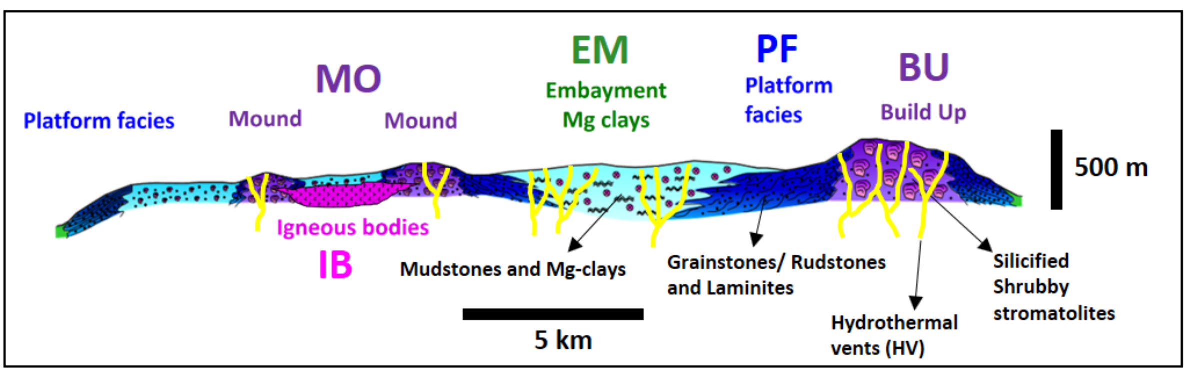

Figure 2.5. Barra Velha Formation main depositional facies at Mero Field (modified from Sartorato et al, 2018). The intrusive igneous bodies are of Santonian age and alkaline signature.

Mero Field key reservoir metrics	
Fluid API	29º
Structural closure	211 km^2
Column height	600 m
Net pay	194-420 m
Porosity	10-14%
Permeability	12-12,000 mD
CO_2 content	~ 44%
Recoverable volume	3.3 billion bbl

Table 2.1 – Key metrics of the Mero Field (after Carlotto and Silva, 2018).

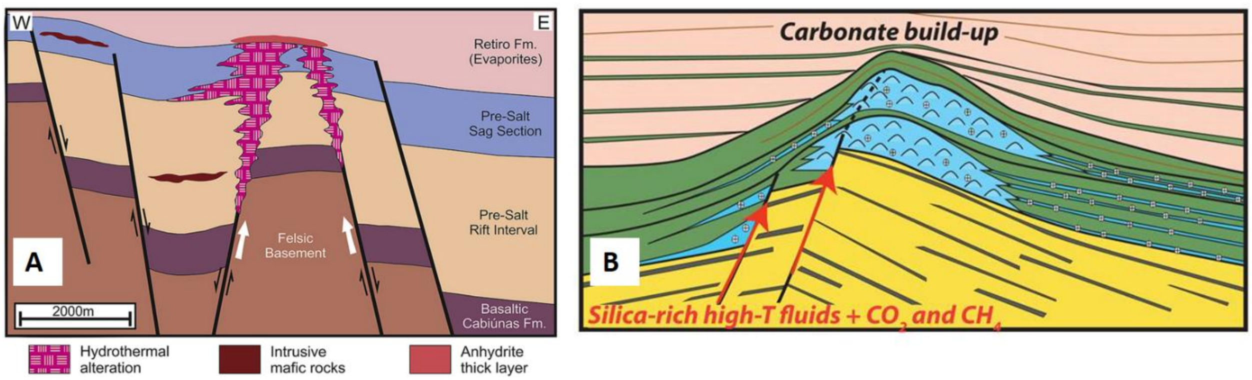

Figure 2.6. A) Proposed model by Lima et al., (2020) for pre-salt reservoirs of Northern Campos Basin. B) Hydrothermal alteration model proposed by Poros et al., (2017) for Kwanza Basin, Angola.

Dataset

We used marine 3D seismic data and reservoir parameters from 13 wells drilled at Mero Field and Central Libra (Figure2. 2) that will be described in the next subsections.

Seismic Data

The seismic dataset comprises a Reverse Time Migration (RTM) full stack, angle stacks, and an interval velocity model obtained by Full Waveform Inversion (FWI). The acquisition is a

3D marine, broadband, narrow azimuth streamer towed survey acquired in 2012 (before any test or production had started) and reprocessed in 2016 (Figure 2.2). Due to its size (17 GB), the original dataset was cropped around the study area (Figure 2.2), and the final volume is 2 GB for each calculated seismic attribute and the velocity model (Figure 2.7).

Seismic inversion data (both acoustic and elastic) provides one of the best means for facies discrimination and decameter flow unit identification (Penna and Lupinacci, 2020). However, such data are rarely available during early exploration phases (wildcat and appraisal drilling). When available, they may be unreliable due to lack of wells for more accurate wavelet estimation. As the main objective is to identify reservoir sweet spot as soon as possible in the E&P project timeline, in this work, we focus on the classical attributes described by Chopra and Marfurt (2007). They can be calculated in AASPI (Attribute Assisted Seismic Interpretation- http://mcee.ou.edu/aaspi/) and also in almost all of the seismic interpretation commercial software by using a full stack amplitude volume as input. In addition to these classical attributes, we also used as input the interval velocity model obtained via full waveform inversion (FWI), using higher frequency models (Castiello et al. ,2022), and Fatti's AVO (Fatti et al., 1994) volumes using angle stacks. This seismic input suite is available for most modern exploration projects, even in frontier areas. Although more computationally expensive than the post stack attributes, these seismic velocities and AVO volumes commonly form part of modern deep water exploration workflows.

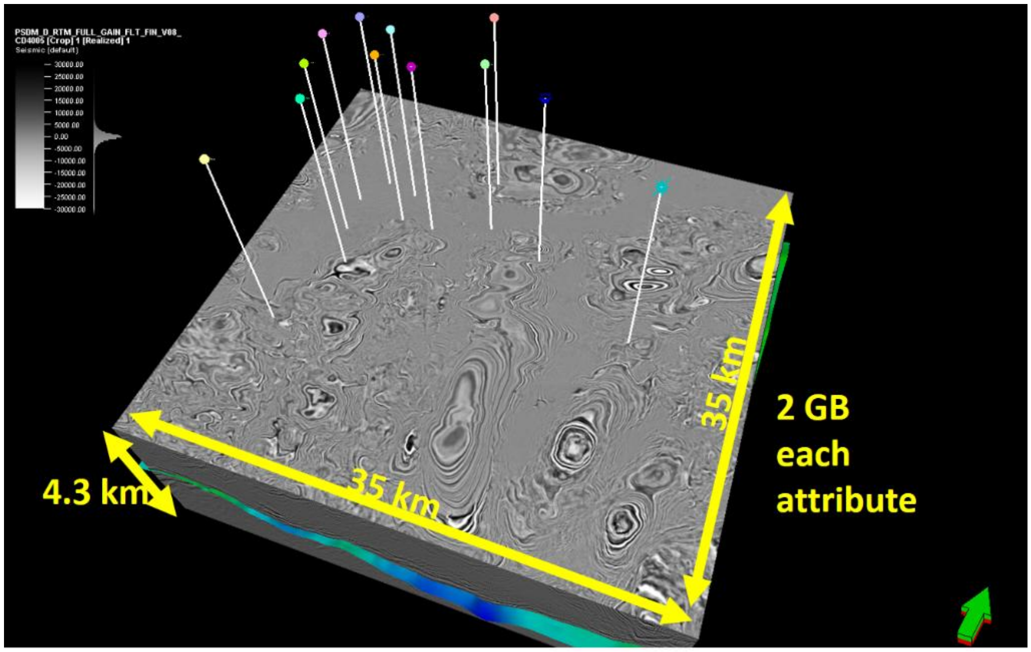

Figure 2.7. 3D view of the input seismic cube after cropping around the area of interest.

A Seismic Acquisition Parameters
Santos Ph1 6A– CGG- 2011/12

Area	2,848 km^2
Number of cables	8-10
Maximum offset	8,000 m
Channels/cable	640
Shot interval	25 m (flip-flop)
Station interval	12.5 m
Shot line interval	500 m
Cable interval	100 m
Sample interval	2 ms
Bin size	6.25 x 25 m
Cable depth	9 m
Source depth	6 m
Azimuth	178/358
Nominal fold	80
Traces/km^2	512,000

B 35 Hz RTM reprocessed data
(CGG-2016)- Cropped version

Area	1,205 km^2
Number of inlines	1,451
Number of crosslines	1,365
Bin size	25 x 25 m
Sample interval	5 m
Samples/trace	861
Total number of traces	1,980,615
Total number of voxels	1,705,309,515
Volume size	2 GB

Table 2.2 – A) Seismic acquisition parameters; and B) Geometry of the cropped version of the velocity and reverse time migrated depth volumes.

Well data

We used data from one wildcat and twelve appraisal wells drilled between 2010 and 2018. Wells A to H have both electric logs, rock and fluid samples and drill stem test data (DST), a temporary well completion technique for reservoir productivity data acquisition by measuring the reservoir pressure behavior during oil production and static periods (Bourdet, 2002). Wells X1, X2, X3, X4, and X5 have only electric logs and rock/fluid samples. Wells A (injector), C, H, and X3 (producers) were part of an anticipated production system (long term DST), during which they were put in production from 2017 to 2021 (44 months) for field scale measurements of productivity parameters (Figure 2.8).

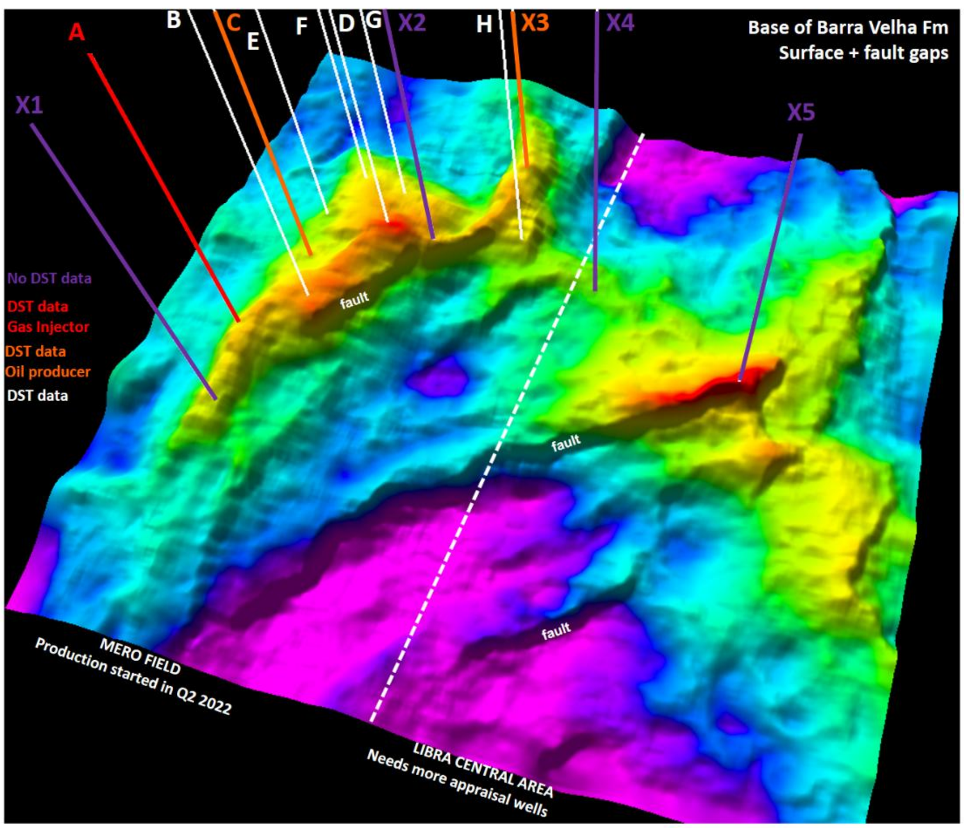

Figure 2.8. 3D view of the base of the Barra Velha reservoir surface (and fault gaps) at Mero Field (left) and Central Libra Appraisal Plan area (right) showing the wells used in this study.

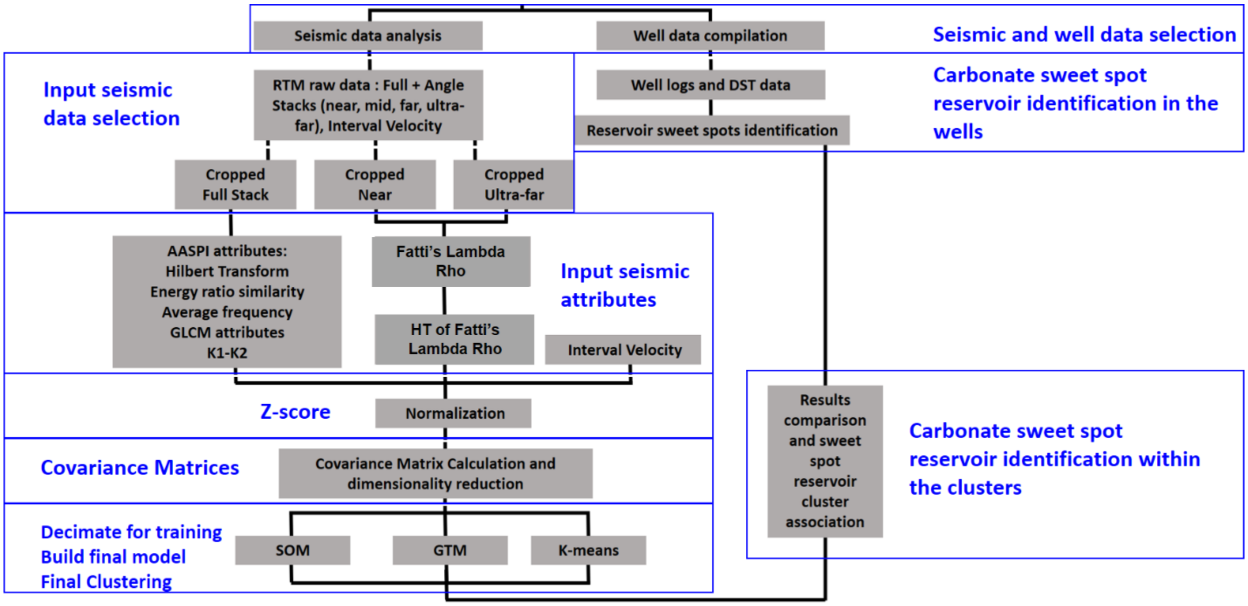

Figure 2.9. Workflow used in this study. (RTM = Reverse Time Migration, DST = Drill Stem Test), GLCM = Gray Level Co-occurrence Matrix, K1 and K2 = maximum positive and negative curvatures, HT = Hilbert Transform, SOM = Self Organizing Maps, GTM = Generative Topographic Models).

Method

Seismic data conditioning, interpretation and attribute extraction

Figure 2.9 shows the workflow applied in this study, and each step is detailed below. Each of the seismic data volumes described previously is 17 GB. For this reason, they were cropped about the area of interest (AOI) in the Mero Field and Central Libra Appraisal Plan area (Figures 2.2 and 2.7). After cropping, each seismic volume became 2 GB and all of them were analyzed and interpreted in the depth domain. Table 2.2A summarizes the seismic acquisition parameters, and Table 2.2B describes geometry of the cropped seismic dataset. After seismic-well correlation, the top and base of the main reservoir unit (Barra Velha-BVE) were interpreted in the full stack

RTM volume. From this input volume, for the BVE reservoir, we calculated the following attributes which are thoroughly described in Chopra and Marfurt (2007):

- Hilbert transform of the amplitude (HTAM): it is a complex trace attribute that mimics a seismic bandwidth acoustic impedance by ninety degrees phase rotation of the input amplitude volume (Taner et al., 1979). Any other kind of relative acoustic impedance attribute could be used instead.
- Energy Ratio Similarity (ERS): it is a geometric attribute based on the eigenstructure of the covariance matrix formed from the traces in the analysis cube. It is good for identification of faults, fractures, and cavities (Gersztenkorn and Marfurt,1999).
- Average frequency (AVF): it is another complex trace attribute computed using the Hilbert transform. The frequency values are close to the dominant frequency at the strongest reflection events (Taner et al., 1979). Lower frequency response can be attributed to intrinsic attenuation associated with higher permeability and geometric attenuation due to scattering related to fracturing and karstification (Zhang et al., 2008).
- Gray Level Co-occurrence matrix (GLCM):it is a voxel-based texture attribute good for faults and fractures detection that has three main types (Haralick et al., 1973 and Matos et al., 2011):
    - GLCM Energy (GLEN): measures the uniformity of lateral amplitude response.
    - GLCM Entropy (GLET): measures the randomness of lateral amplitude response.

    - GLCM Homogeneity (GLEH): measures overall smoothness of an image. Highly contrasting voxels lead to low homogeneity values. For this reason, it is also good for reflector continuity assessment.
- Maximum positive (K1) and negative (K2) curvatures: they are measurements of reflectors convexity (K1) and concavity (K2) by fitting circles' curvatures. The reflectors' curvatures are inversely proportional to the radius of the mentioned circles (Chopra and Marfurt, 2007).

Additionally, we used a velocity model and an amplitude versus offset (AVO) attribute as inputs:

- Interval velocity (VEL): It is the velocity model built via FWI during seismic processing. These velocity models have more frequency content than those obtained with tomography (Castiello et al.,2022)
- Hilbert transform of Fatti's Lambda-rho attribute (HTLR): It was obtained after computing the A and D terms of Fatti's approximation of Zoepritz's equation by using near and far angle stacks. According to Fatti et al (1994), the sum 2A + D is proportional to the sum of P and S reflectivity. When the Hilbert transform is applied to the Fatti's Lambda-rho attribute, it mimics a seismic bandwidth P impedance minus S impedance volume (Ip-Is).

Well data compilation

For reservoir quality assessment, we used petrophysical totalization reports which contain the main reservoir parameters (porosity, gross pay, net pay, fluid type and water saturation). We also use the lithogeochemical spectroscopy log to identify hydrothermal zones identification by measuring relative concentration of major chemical elements (Ca, Fe, Mg, Si, H and Cl) (Griffiths, 2010). In wells with DST data, we compiled data of productivity index (flow rate divided by pressure drop times skin factor) and flow capacity (hydraulic transmissibility times fluid viscosity, known as *kh*). These two are the main parameters used for decision making for future production wells completion (Bourdet, 2002). For wells used in the anticipated production system, plateau oil production flow rates and gas injection rates values were compiled. In Table 2.3, we have all reservoir quality parameters for the BVE reservoirs at the Mero Field and Central Libra appraisal plan area.

Well	Facies	Main Lithologies	Mean Porosity (%)	Gross pay (m)	Net pay (m)	Net to gross (%)	Silicon anomaly (> 50%)	Productivity Index (m^3/d)/(kgf/cm^2)	Flow Capacity (mD.m)	Oil production plateau (bbl/d)	Gas Injection plateau (m^3/d)
A	Build up	Silicified Shrubby stromatolites	12.95	286	276.63	96.7	yes	2,200	350,000	-	75,000
B	Platform	Spherulitites, Grainstones and laminites	11.3	109	102.39	94	no	-	-	-	-
C	Platform	Spherulitites, Grainstones and laminites	9.8	311	199	60	no	9.25 upper BVE 367 lower BVE	3,230 upper BVE 50,000 lower BVE	40,000	-
D	Platform	Grainstones and laminites	12.35	57.61	50.06	87	no	73	6,500	-	-
E	Platform	Spherulitites, Grainstones and laminites	12.7	244.7	146.8	60	no	-	-	-	-
F	Platform	Spherulitites, Grainstones and laminites	11.3	178.7	150.7	84	no	14	2,235	-	-
G	Platform	Spherulitites, Grainstones and laminites	10.85	219.64	191.26	87	no	132.66	24,000	-	-
H	Platform	Spherulitites, Grainstones and laminites	9.65	389	187.72	48.26	no	55.8	8,130	20,000	-
X1	Build up	Silicified Shrubby stromatolites	13	410	398.53	97.2	yes	-	-	-	-
X2	Build up	Locally Silicified Shrubby stromatolites and travertines	10.8	551.92	313.67	57	yes	-	-	-	-
X3	Build up	Silicified Shrubby stromatolites	14.2	383.46	304.95	79.5	yes	-	-	48,000	-
X4	Embayment	Spherulitites and mudstones	<5	0	0	0	no	-	-	-	-
X5	Build up	Silicified Shrubby stromatolites	10.4	382.2	167.69	44	yes	-	-	-	-

Table 2.3 – Reservoir quality data compilation for wells (wildcat and appraisal) drilled at Mero Field and Central Libra appraisal plan area for the Barra Velha Formation (BVE). Wells in blue are those drilled in the sweet spot reservoirs: silicified shrubby stromatolites within build up facies.

Carbonate facies and related seismic attributes

The reference reservoir facies model adopted for the Mero Field in this work is the one proposed by Sartorato et al. (2018) and modified in Figure 2.5. It is a well data based (logs and rock samples) simplistic but intuitive model that shows the distribution of the main facies observed at Barra Velha Formation across the field: build ups, mounds, platform facies, hydrothermal vents, embayment facies, and Santonian age igneous intrusions. The main reason for choosing the attributes described above is their availability at all phases of any E&P project and the very low correlation between the chosen attributes shown by their correlation matrix (Table 2.4). The lower the correlation among these, the higher the individual contribution of the attributes for facies discrimination (non-redundant attributes). For example, as the GLCM attributes have very high correlation among them (redundant attributes), a test with SOM after removing both entropy and homogeneity was undertaken. Based on the authors' experience in carbonate facies seismic identification (Maas et al., 2019), each facies were associated with a seismic signature observed at each one of the seismic attributes and are summarized in Table 2.5.

In Figure 2.10, we can observe different attribute signatures of the carbonate facies. I focus here on the main features of interest: silicified carbonate build ups and mounds (main producers sweet spot reservoirs) and igneous intrusions (they act as drilling and flow barriers). The carbonate build ups (BU) and mounds (MO) are often characterized by:

- negative HT anomalies (amplitude and Fatti´s Lambda-Rho), due to acoustic impedance drop after hydrothermal silica precipitation (lower density than calcite).
- low energy ratio similarity, GLCM energy and homogeneity anomalies related to faults/fractures/karstification.

- high K1 anomalies related to its convex up geometry.

The igneous intrusions (IB) have the following signatures:

- positive HT anomalies (amplitude and Fatti´s Lambda-Rho), due to their intrinsic high acoustic impedance.
- high energy ratio similarity, GLCM energy and homogeneity anomalies related to their massiveness, continuity and lack of faults, fractures and caves.
- near zero K1-K2 values related to its flat geometry.

Attribute	ERS	GLCM Energy	GLCM Entropy	GLCM Homogeneity	HT of Fatti's Lambda-Rho	HT of amplitude	Average frequency	K1	K2	Velocity
Energy Ratio Similarity	1.00	-	-	-	-	-	-	-	-	-
GLCM Energy	0.37	1.00	-	-	-	-	-	-	-	-
GLCM Entropy	-0.42	-0.98	1.00	-	-	-	-	-	-	-
GLCM Homogeneity	0.48	0.92	-0.97	1.00	-	-	-	-	-	-
HT Fatti's Lambda- Rho	0.01	0.01	-0.02	0.02	1.00	-	-	-	-	-
HT of amplitude	-0.01	0.02	-0.01	-0.01	0.68	1.00	-	-	-	-
Average frequency	-0.03	0.04	0.08	-0.01	-0.09	-0.13	1.00	-	-	-
K1	-0.13	-0.12	0.13	-0.13	0.00	-0.05	0.00	1.00	-	-
K2	-0.02	-0.01	0.01	-0.01	0.00	0.00	0.00	0.48	1.00	-
Velocity	0.04	0.03	-0.04	0.04	0.03	0.02	-0.05	-0.04	0.00	1.00

Table 2.4 – Correlation matrix of selected attributes within the Barra Velha Formation at the studied area. The closer to 1 or -1, the higher the correlation. The values shown in red are the high correlation among the redundant GLCM´s attributes.

Attribute	Facies	Signature	Geological reason
HTAM Hilbert Transform of input Amplitude: Mimics Acoustic Impedance	Build ups (BU) & mounds (MO)	Low	Higher porosity related to in situ facies (boundstones-shrubby stromatolites) and intense fracturing/karstification. Hydrothermal silicification reduces bulk density.
	Platform Facies (PF)	Medium	Lower porosity and less fractures/karstification
	Embayment (EM)	Low	Low porosity-permeability mudstones
	Igneous bodies intrusions (IB)	High	Zero porosity and almost no fracturing.
HTLR Hilbert Transform of Fatti´s Lambda-Rho: Mimics Ip-Is	Build ups (BU) & mounds (MO)	Low	Higher porosity related to in situ facies (boundstones-shrubby stromatolites) and intense fracturing/karstification. Hydrothermal silicification reduces bulk density.
	Platform Facies (PF)	Medium	Lower porosity and less fractures/karstification
	Embayment (EM)	Low	Low porosity-permeability mudstones
	Igneous bodies intrusions (IB)	High	Zero porosity and almost no fracturing.
VEL Interval Velocity: built via FWI	Build ups (BU) & mounds (MO)	Low	Higher porosity related to in situ facies (boundstones-shrubby stromatolites) and intense fracturing/karstification. Hydrothermal silicification reduces bulk density.
	Platform Facies (PF)	Medium	Lower porosity and less fractures/karstification
	Embayment (EM)	Low	Low porosity-permeability mudstones
	Igneous bodies intrusions (IB)	High	Zero porosity and almost no fracturing.
ERS Energy Ratio Similarity: Reflector Continuity (Faults, Fractures, Karst)	Build ups (BU) & mounds (MO)	Low	Intense fracturing /karstification
	Platform Facies (PF)	High	Lower fracturing /karstification
	Embayment (EM)	High	No fracturing/karstification
	Igneous bodies intrusions (IB)	High	No fracturing /karstification
GLEN GLCM Energy: Pixel based texture: Continuity	Build ups (BU) & mounds (MO)	Low	Intense fracturing /karstification
	Platform Facies (PF)	Medium	Lower fracturing /karstification
	Embayment (EM)	Medium	No fracturing /karstification
	Igneous bodies intrusions (IB)	High	No fracturing /karstification
GLET GLCM Entropy: Pixel based texture: Reflector disorganization	Build ups (BU) & mounds (MO)	High	Intense fracturing /karstification
	Platform Facies (PF)	Medium	Lower fracturing /karstification
	Embayment (EM)	Medium	No fracturing /karstification
	Igneous bodies intrusions (IB)	Low	No fracturing /karstification
GLHO GLCM Homogeneity: Pixel based texture: Smoothness	Build ups (BU) & mounds (MO)	Low	Intense fracturing /karstification
	Platform Facies (PF)	Medium	Lower fracturing /karstification
	Embayment (EM)	Medium	No fracturing /karstification
	Igneous bodies intrusions (IB)	High	No fracturing /karstification
AVF Average Frequency : attenuation measurement	Build ups (BU) & mounds (MO)	Low	Attenuation due to higher porosity and fracturing/karstfication
	Platform Facies (PF)	Medium	Lower porosity and lower fracturing/karstfication
	Embayment (EM)	High	Lower porosity and no fracturing/karstfication
	Igneous bodies intrusions (IB)	Low	Zero porosity and no fracturing/karstfication
K1 Maximum Positive Curvature: Reflector convexity	Build ups (BU) & mounds (MO)	Positive	Top of these features are convex zones
	Platform Facies (PF)	Near Zero	Top of these features are more flat zones
	Embayment (EM)	Low to negative	Top of these features are concave zones
	Igneous bodies intrusions (IB)	Near Zero	Top of these features are more flat zones
K2 Maximum Negative Curvature: Reflector Concavity	Build ups (BU) & mounds (ME)	Low to positive	Top of these features are convex zones
	Platform Facies (PF)	Near Zero	Top of these features are more flat zones
	Embayment (EM)	Negative	Top of these features are concave zones
	Igneous bodies intrusions (IB)	Near Zero	Top of these features are more flat zones

Table 2.5 – Attributes signatures related to each facies of the Barra Velha Formation at the studied area.

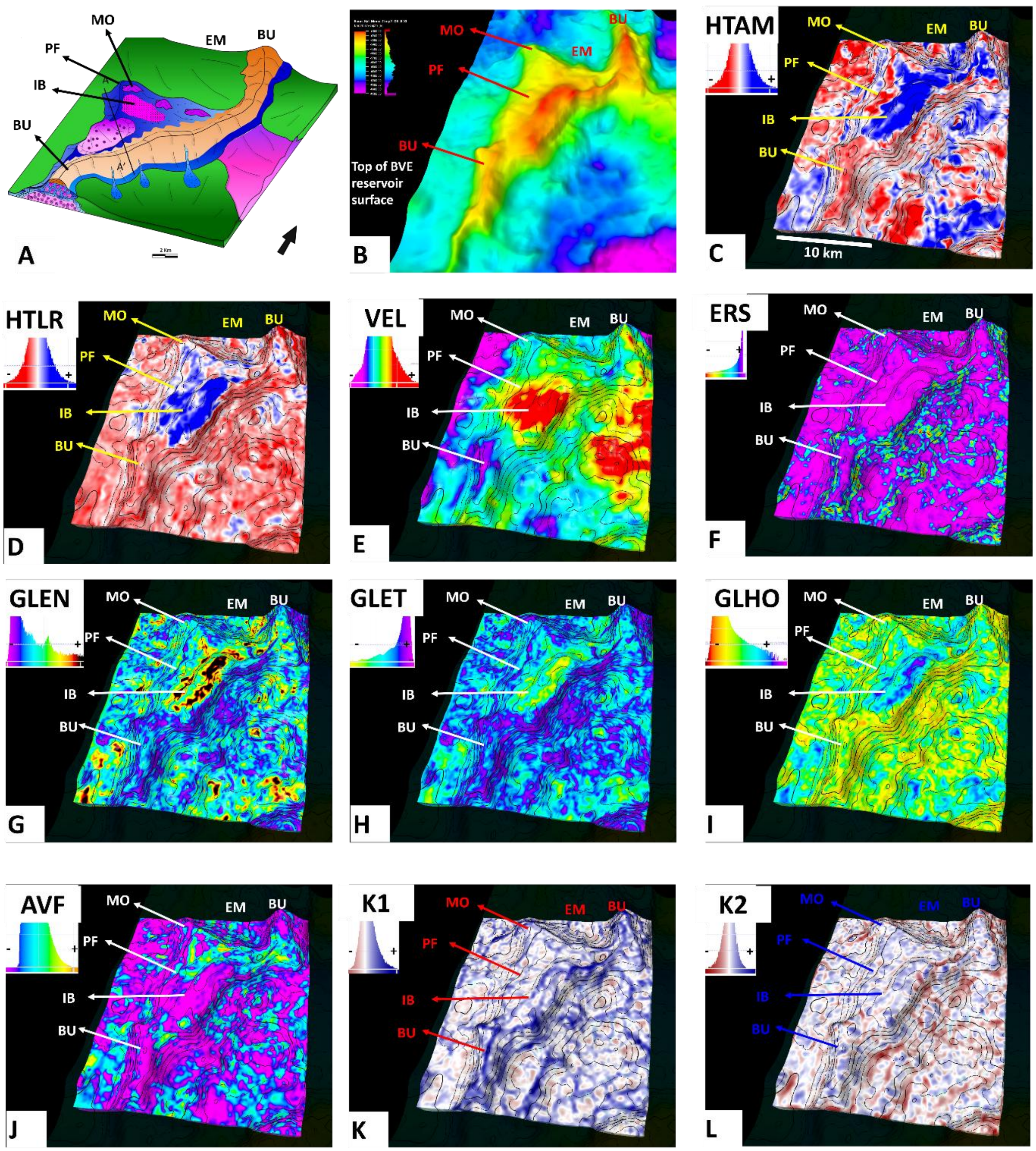

Figure 2.10. 3D view of different seismic attributes used in this study. All of them were extracted within the Barra Velha Formation using its top and bottom horizons. A) Conceptual depositional model for the Barra Velha Formation of the Mero Field proposed by Sartorato et al., (2018). (BU= Build Up, PF= Platform facies, MO= Mounds, EM= Embayment facies, IB= Igneous bodies); B) Depth map of top of Barra Velha surface; C) Hilbert Transform of Amplitude; D) Hilbert Transform of Fatti´s Lambda-Rho; E) Interval velocity; F) Energy Ratio Similarity; G) GLCM Energy; H) GLCM Entropy; I) GLCM Homogeneity; J) Average frequency; K) Maximum positive curvature; L) Maximum negative curvature.

Unsupervised clustering basic theory

Among the unsupervised clustering techniques, there are some that can cope with seismic multi-attributes. They are cross plotted in a **n**-D space where **n** is the number of input attributes and, thus, detect clusters with similar attributes responses and subtle variations that are not perceptible for human interpreters (La Marca, 2020). As it is unsupervised, nothing but normalization is done with the input attributes, and the output models are then interpreted according to the geology of the area. The theoretical aspects of each clustering technique will not be detailed here. A very good explanation of SOM can be found in the works of Kohonen (1982), Matos et al. (2010), and La Marca (2020). Zhao et al., (2015) have done a good explanation on GTM and K-means clustering techniques.

After the outlier removal, each input attribute is normalized to avoid problems with different ranges of the inputs. The normalization method used here was z-score, in which all the inputs have a distribution with mean equal zero and standard deviation equal one. After this, covariance matrices were computed, normalized inputs were cross plotted in a high dimensional space (one dimension per attribute) and distances from the prototype vectors to each data point were calculated. In K-means, clustering is performed using the mean values of each one of the optimal number of clusters (in this case, eight). In SOM, this is done by competition of the prototype vectors when their distances to the data points are calculated. In GTM, the distances are better calculated using probability density functions. The final clusters are then projected onto a lower dimensional manifold using the eigenvectors (Zhao et al, 2015). In AASPI, both SOM and GTM algorithms calculate a maximum number of 256 clusters after projection onto the latent space (Figure 2.11). This leads to a high capability of subtle variations identification among the clusters.

The interpreter can use his geological reasoning to merge the output clusters according to a priori knowledge. In the K-means algorithm, the number of clusters is limited to a much lower value (eight in our case after testing a range between four and twelve). This led to some limitations that will be further detailed.

For comparison purposes, unsupervised clustering was performed using three different methods and two attribute combinations:

1) SOM1, GTM, and K-means (six attributes): using Hilbert transform of Fatti´s Lambda-Rho, interval velocity, energy ratio similarity, GLCM energy, and K1-K2 curvatures.

2) SOM2 (eight attributes): using Hilbert transform of amplitude, average frequency, energy ratio similarity, GLCM (energy-entropy-homogeneity), and K1-K2 curvatures.

The objective of this comparison is to evaluate which clustering technique produces a more geologically reasonable result and the impact of different attribute number and combinations on the clustering performance in terms of computation time.

The final clustered volumes (four models) were converted to SEG-Y format and exported to a commercial seismic interpretation software, where vertical sections and 3D opacity techniques were applied to better visualize the features of interest as the most productive carbonate buildups (and non-drilled similar structures which can be considered as appraisal candidates) and intrusive igneous bodies.

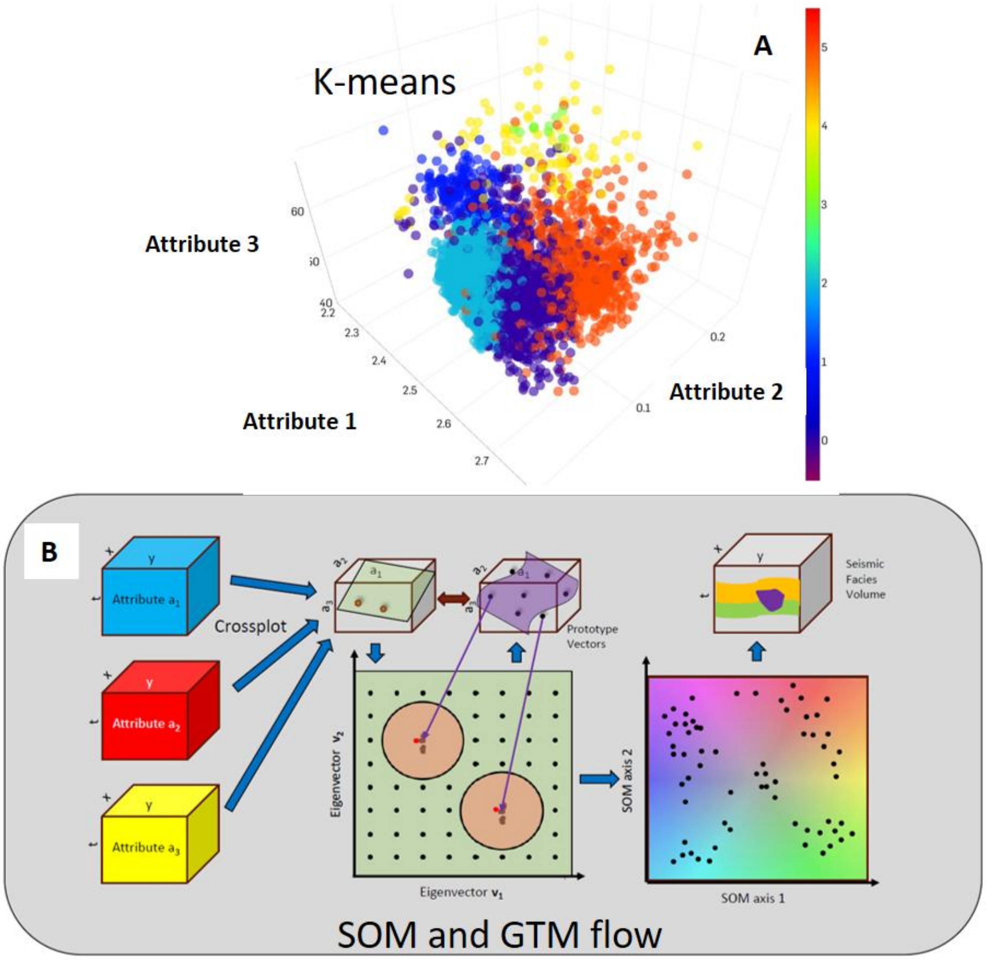

Figure 2.11. A) Hyperdimensional cross plotting: The number of dimensions of clustering is equal to the number of input attributes, in our case from 6 to 8. B) illustration of the basic theory of the clustering algorithms. SOM= Self Organizing Maps, GTM= Generative Topographic Maps. (Modified from Zhao, 2015).

Results

In terms of computational time for clustering, SOM and GTM took 1.5 hours and K-means took 0.35 hours using 32 processors of a 64 GB RAM, 2.8 GHz central processing unit.

Figure 2.12 shows the four different models generated by the clustering algorithms with different attribute combinations. These models are compatible with the published ones made by human interpreters (Figures 2.5 and 2.6) but with the advantage that they took much less time to be built. In Figure 2.13A, we have a structural map of the top of BVE reservoir with the studied wells and one of the identified leads (possible appraisal candidates). It also shows the paths of the vertical sections of Figure 2.14. In Figure 2.13B, the 2D color bar of the four final clustering models are shown.

Vertical sections

Figure 2.14 shows four vertical sections passing through three of the studied wells (with their composite logs) and one of the identified leads. The model SOM2 was omitted due to page space limitations and its high similarity to the SOM1 model. For each one of the wells and the lead, the figure shows the seismic input (RTM full stack), SOM1, K-means, and GTM models. At the Well C location, all of the clustering algorithms identified the igneous body at the BVE reservoir top. At this well, BVE is represented by platform facies mainly composed of spherulitites, grainstones and laminites with medium to high reservoir quality parameters (Table 2.3). These platform facies (PF) were better highlighted by the GTM model probably due to its probabilistic approach of clustering. In SOM1 model, they are less evident, but the clusters are more prone to interpreted as platform facies. In K-means model, both platform end embayment facies (EM) related clusters were merged into one, which represents a limitation due to the much lower number of clusters (eight) of the K-means algorithm compared to the 256 clusters of both SOM and GTM. But even with this drawback, the igneous bodies were clearly detected by the K-means model. Nearby Well C location, mounds (MO) and hydrothermal vents (HV) features can be observed. This can be an explanation for the above average productivity parameters (both productivity index and flow capacity compiled at Table 2.3) when compared to other platform facies wells (Wells D, F, G, and H). At Well X4 location, the worst well in the area in terms of reservoir quality, all models detected the very low porosity and permeability mudstones of the embayment facies (EM) (non-reservoirs) intruded by the igneous bodies (IB). At Well X3, the most productive well drilled so far in the Mero Field, the porous and permeable carbonates of BVE (Table 2.3) related to shrubby stromatolitic build ups are strongly silicified and karstified due to a pervasive hydrothermal event which is well documented by Petrobras internal reports. Due to net porosity gain, this hydrothermal silicification is responsible for the excellent reservoir parameters and productivity shown at Table 2.3. At this well location, all the four models detected the build up feature. At the lead location at the Central Libra Area (close to Well X5), all clustering models identified a likely carbonate build up feature very similar to that drilled by wells A, X1, X2, X3, and X5. For this reason, it is as a good candidate for future drilling.

3D view of the clustering models

Next, geobodies of the main features of interest were extracted: the carbonate sweet spot reservoirs (mainly build ups and mounds) and the intrusive igneous bodies. Secondarily, platform facies and hydrothermal vents were also highlighted. An opacity mask was applied for each model, through which only desired clusters related to the interpreted feature are visible. In Figure 2.15, we can observe the four final clustering 3D models with opacity to the best feature detected by each model. In Figure 2.15A, the SOM1 model detected with a high degree of precision the Santonian intrusive igneous bodies (pink geobodies- mostly sills). The wells B, C, D, X3, and X4 have penetrated those sills. In Figures 2.15B and 2.15D, we can note the geobodies of sweet spot reservoirs, mainly the build ups and mounds composed of highly fractured, karstified and silicified shrubby stromatolites. Secondarily, we can see, mainly in the structural lows, subvertical dyke-like features that are interpreted as hydrothermal vents that are probably responsible for the pervasive silicification observed in this area. In Figure 2.15C, we applied opacity to both buildups, mounds, hydrothermal vents and platform facies-related clusters. In this 3D view, we can see at the structural highs (above the oil water contact- OWC), the lateral continuity of the platform facies connecting the build ups and mounds, this is an explanation to the hydraulic connection observed at Mero field and its disconnection related to Central Libra Area. In Figures 2.15B, 2.15C, and 2.15D, we can observe undrilled features that are very similar to those penetrated by highly productive wells (X3 and C) and very high-quality BVE reservoirs wells (A, X1, X2, X3, and X5). The features identified around Well X5 are likely thick (> 300 m) carbonate build ups with very high porosity, permeability and also very high productivity index and flow capacity values.

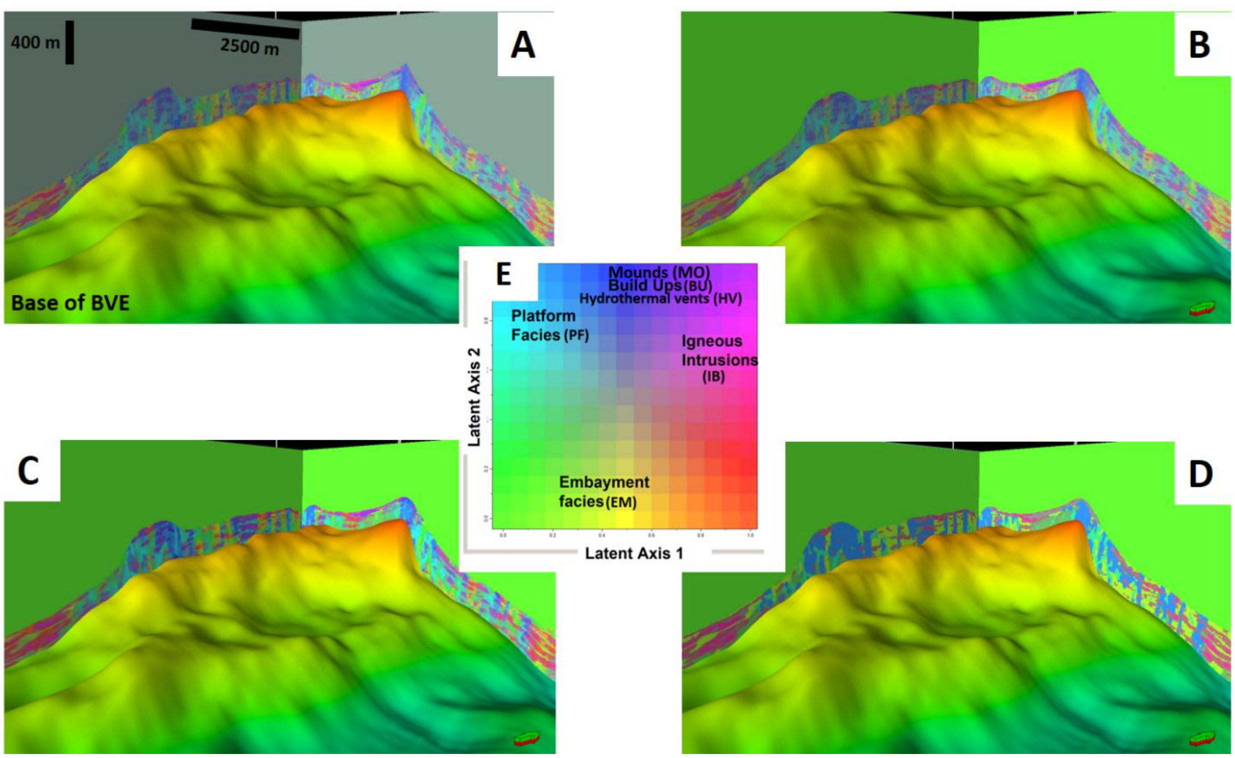

Figure 2.12. 3D view (Inline, Crossline and base of Barra Velha reservoir surface) of four final clustering models for BVE reservoir at the studied area. A) SOM1 using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. B) SOM2 using eight attributes: HTAM, AVF, ERS, GLEN, GLET, GLHO, K1, and K2. C) GTM using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. D) K-means using six attributes: HTLR, VEL, ERS, GLEN, K1, and K2. E) 2D color bar of the unsupervised clustering models, use this color bar when examining Figure2.14.

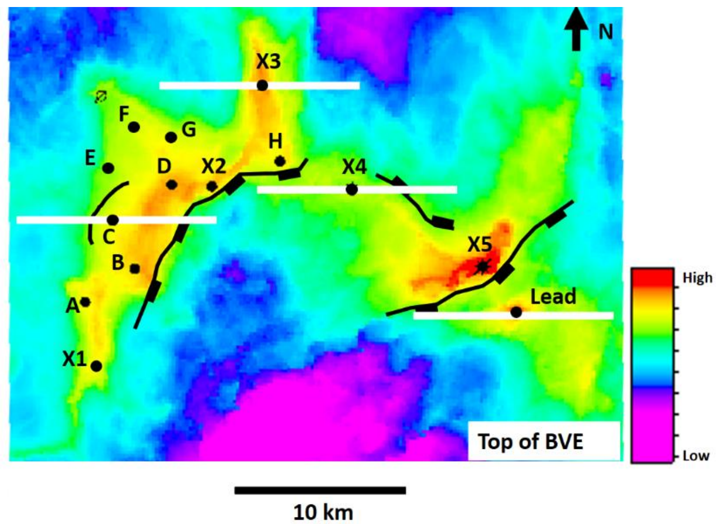

Figure 2.13. Seismic structural map of top of Barra Velha (BVE) reservoir with the studied wells and one lead. The horizontal white lines are the vertical sections paths shown in Figure 2.14.

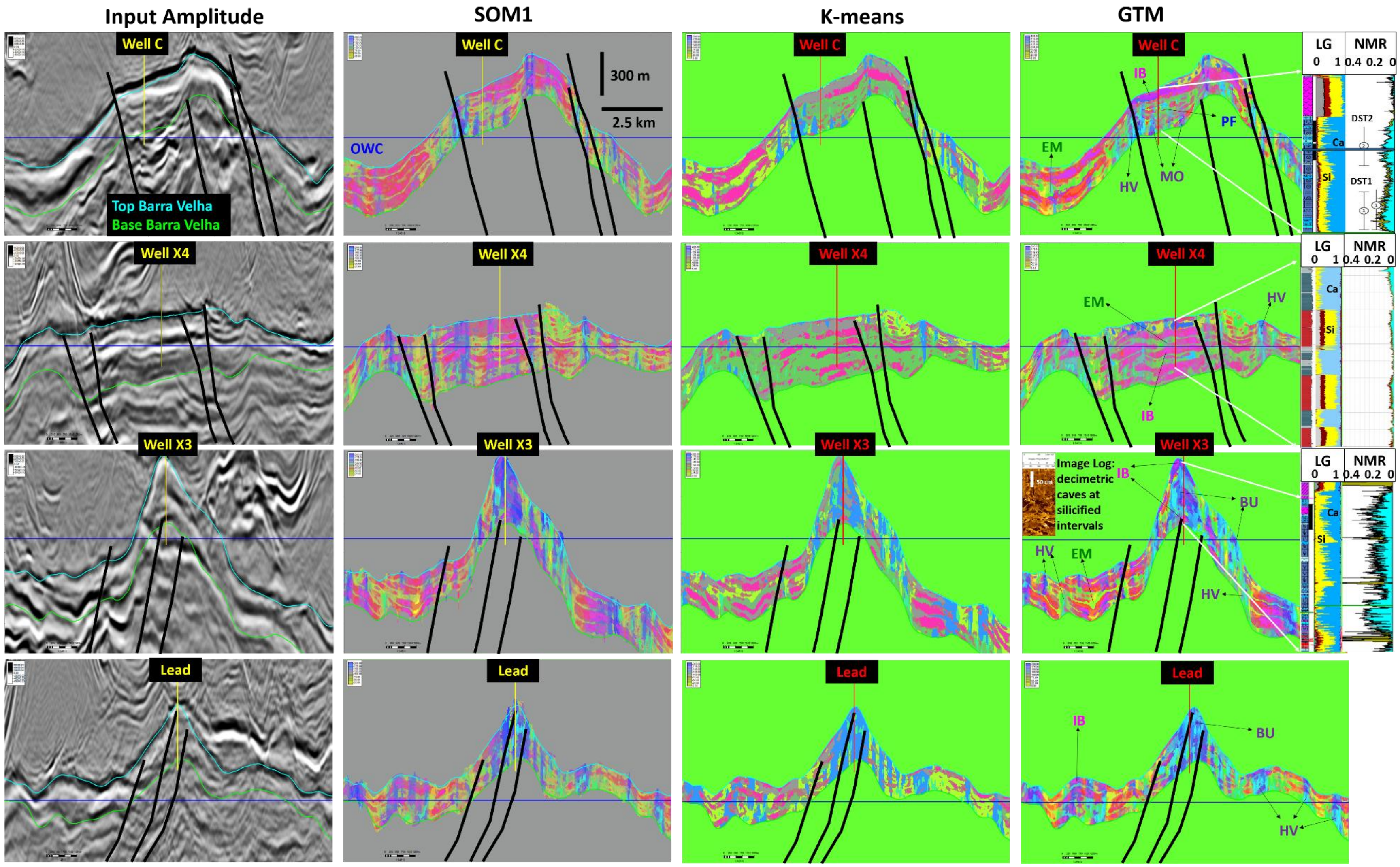

Figure 2.14. East-West vertical sections passing through some of the studied wells and one lead. The input amplitude, SOM1, K-means and GTM models are shown. In GTM model vertical sections we interpreted the main features: build ups (BU), mounds (MO), platform facies (PF), igneous bodies (IB), embayment facies (EM) and hydrothermal vents (HV). In the well logs close to GTM models vertical sections, we have litogeochemical logs (LG) with the highlighted silicon (Si) and calcium (Ca) relative concentrations, and nuclear magnetic resonance (NMR) porosity log. In Well C, the silicon content is related to clay minerals in the platform facies limestones. In Well X4, the silicon content is related both to clay minerals in the embayment facies and silicate minerals in the igneous bodies. In Well X3, the silicon content is related to a pervasive hydrothermal silicification amplifying the metric cave porous system, which is shown in the image log. OWC = oil water contact.

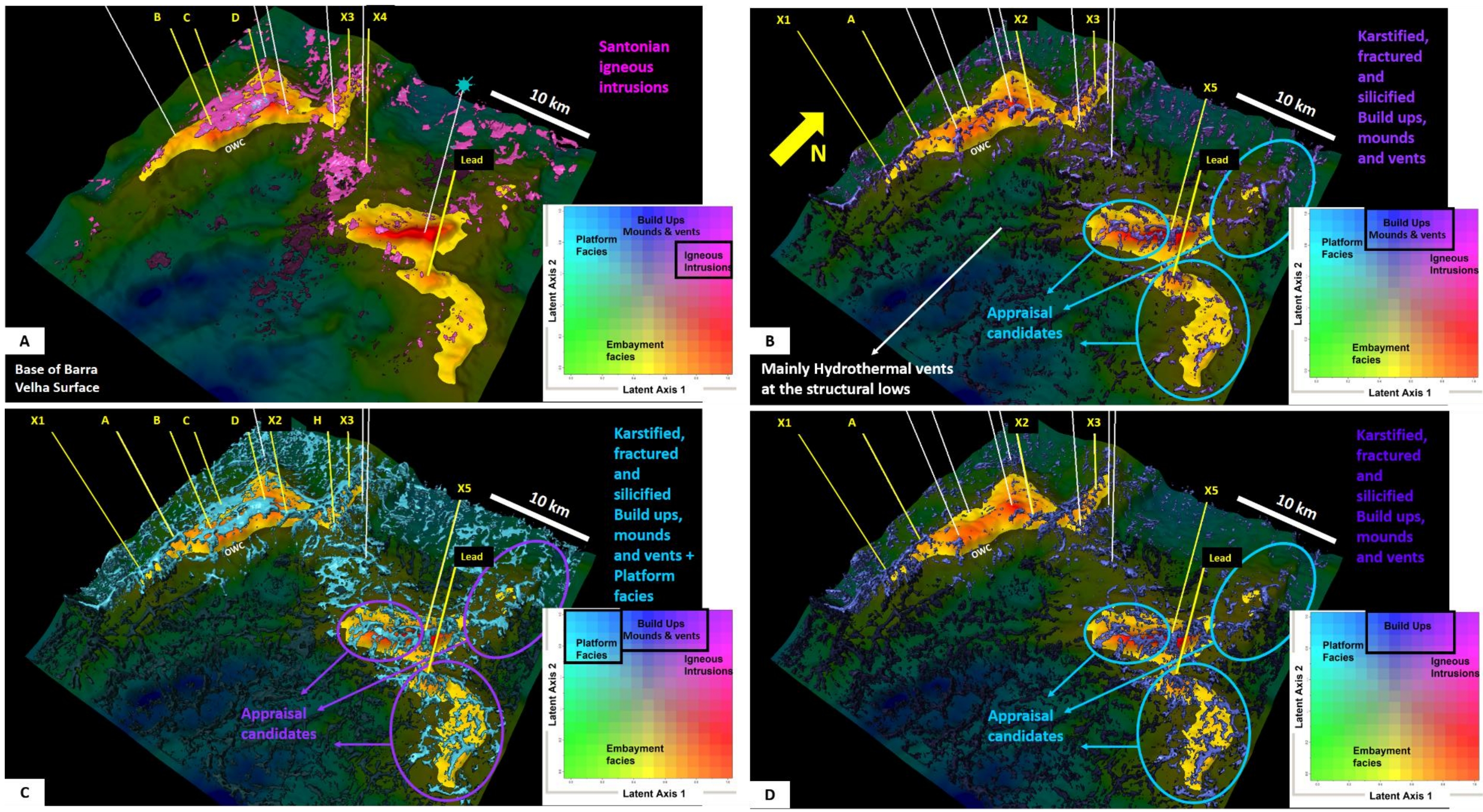

Figure 2.15. 3D view of the clustering models using opacity to highlight the main features of interest over the base of Barra Velha reservoir surface and the Oil Water Contact (OWC) plane. A) SOM1 model with opacity to the clusters that are interpreted as sill intrusions. Wells highlighted in yellow penetrated those sills. B) SOM2 model with opacity to the clusters interpreted as sweet spot reservoirs: highly silicified and fractured build ups, mounds and hydrothermal vents. Wells highlighted in yellow penetrated those excellent reservoirs. C) GTM model with opacity to the clusters interpreted as both build ups, mounds, hydrothermal vents and the platform facies. Wells highlighted in yellow penetrated those reservoirs. D)K-means model with opacity to the clusters interpreted as sweet spot reservoirs: highly silicified and fractured build ups, mounds and hydrothermal vents. Wells highlighted in yellow penetrated those excellent reservoirs.

As they were not used for any training, the wells are very good blind tests. When they are plotted over the clustering results, one can realize that the build ups and mounds sweet spot reservoirs (excellent producers), platform facies (mild producers), sill intrusions and embayment facies (non-reservoirs) are coincident with distinct clusters that are mappable in three dimensions.

Discussion

The attribute combination used in this work has a high geological significance as most of them have a high individual contribution for facies discrimination (very low correlation between attributes). Furthermore, as seen in Figures 2.12 and 2.14, at each clustering model, we can see the signature of certain attributes. For example, at the build ups and their vicinity, it is clearly seen that vertical features are highlighted by the ERS attribute, which has great impact on the delineation of the build ups, mounds, and hydrothermal vents due to their high degree of fracturing, karstification and silicification (Figure 2.10F). In Figure 2.10K, we can see the positive values of K1 aligned over the main build up crests. The combination of these two attributes with the others presented here seems to be enough to avoid ambiguity.

From the four final models, SOM1, SOM2, and GTM are very similar, where GTM showed better facies discrimination probably due to its probabilistic approach during the clustering. The K-means limitations regarding the low number of output clusters affected the discrimination between the platform and embayment facies as their related clusters were often merged, which represents a drawback for this purpose. On the other hand, the build ups, mounds and hydrothermal vents were very well delineated by the K-means algorithm, and it took one third of the computational time. For this reason, K-means seems to be an effective build up detector. The combination of these three clustering techniques was very opportune since it showed that some worked better than others for platform facies identification (GTM). For build ups, mounds and hydrothermal vents all of them worked appropriately. Build up like features were spotted in the NE and SE area of X5 well at Central

Libra Area and could be considered as future targets for appraisal drilling for either oil production or CO_2 injection.

It is very important to mention that seismic identification of most productive reservoirs is vital for project optimization. For example, targets like the Well X3 can produce up to 48,000 barrels per day (bpd). Only four pairs of production-injection wells are required to fulfill the capacity of a 180,000 bpd Floating Production, Storage and Offloading unit (FPSO). For wells like D, E, F, G, and H (that can produce up to 20,000 bpd), nine pairs of wells are necessary to fulfill the capacity of the same FPSO.

The attribute combination used in this work is available for any of the Brazilian pre-salt exploration blocks and their geology is roughly similar to the Mero Field area (mainly for the BVE reservoirs). If working with other geological settings, other attribute combinations should be tried. Because the input attributes are normalized, and covariance is the main metric for high dimensional clustering projections, this methodology can be applied to other seismic surveys. We tested the same attribute combination and SOM clustering in another Petrobras' confidential dataset and the same results were observed: the sweet spot reservoirs drilled and tested by wells were identified within the SOM clusters. So, the same combination can be used in unsupervised clustering for reservoir de-risking in future exploration projects for wildcat and appraisal drilling optimization.

Conclusions

In our unsupervised machine learning analysis of Mero Field, we find that using instantaneous, geometric, interval velocity, and AVO attributes available early in the exploration phase of an oil field, can differentiate highly productive sweet spots and moderately productive platform facies from non-reservoir mudstones and igneous rocks, that form flow barriers and can cause drilling problems. We chose attributes to be not only geologically meaningful in terms of target facies, but also mathematically independent. Given the overall high-quality depth migrated amplitude volume used, our only processing task was some gentle data cleaning and attribute normalization. The proximity of Mero Field to the Central Libra Appraisal Plan, where only two exploratory wells have been drilled to date, showed that the patterns identified at the Mero Field can be readily applied to the Central Libra Area. The SOM, GTM, and K-means clusters from the Central Libra Area are similar to the sweet spots identified at the Mero Field, providing good targets for future drilling, and later in the life of the oilfield, for CO_2 injection. The attribute combination and methodology presented here can be applied to any area with geological similarity for both exploration, appraisal and development optimization.

Acknowledgements

We would like to thank the Libra Consortium (Petrobras, Shell Brasil, TotalEnergies, CNPC, CNOOC and PPSA), the Brazilian National Petroleum Agency (ANP) for releasing the dataset used in this work. Special thanks to Petrobras for sponsoring the first author´s research done so far with the SDA/AASPI Research Group at the University of Oklahoma (OU). Many thanks also to other OU colleagues for suggestions, ideas exchange, and review. We are also very grateful to Dr Kurt J. Marfurt, for revising the manuscript.

References

Bishop C. M., Svensén M., Williams C. K.I., 1998. Developments of the generative topographic mapping, Neurocomputing, **21**, 203-224.

Bourdet D., 2002. Well-test Analysis: The use of advanced interpretation models, Elsevier.

Carlotto M.A, Silva R.C, 2018. Libra: A Newborn Giant in the Brazilian Pre-Salt Province. AAPG Annual Convention and Exhibition, Expanded Abstracts, Salt Lake City.

Carvalho, M. D., and F. L. Fernandes, 2021. Pre-salt depositional system: Sedimentology, diagenesis, and reservoir quality of the Barra Velha Formation, as a result of the Santos Basin tectono-stratigraphic development, *in* Marcio R. Mello, Pinar O. Yilmaz, and Barry J. Katz, eds., The supergiant Lower
Cretaceous pre-salt petroleum systems of the Santos Basin, Brazil: AAPG Memoir 124, p. 121–154.

Castiello, A., Venfield, G., Beenfeldt, J. , Mueller, E. , Khan J. 2022. Integrating FWI and reflection tomography to rejuvenate legacy seismic sata: an example from the Faroes-Shetland Basin: 83rd EAGE Annual Conference & Exhibition, Expanded Abstracts, 1 – 5.

Chopra, S., Marfurt, K.J., 2007. Seismic Attributes for Prospect Identification and Reservoir Characterization: Society of exploration geophysicists.

Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P.J., Levitt, P.R., 1994. Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique: Geophysics, 59, 1362–1376.

Forgey, E., 1965. Cluster analysis of multivariate data: Efficiency vs. interpretability of classification: Biometrics, 21,768.

Gersztenkorn, A., Marfurt, K.J. 1999. Eigenstructure based coherence computations as an aid to 3-D structural and stratigraphic mapping: Geophysics, 64, 1468–1479, doi: 10.1190/1.1444651

Griffiths R., 2010. Ecoscope User’s Guide. Schlumberger.

Haralick, R. M., K. Shanmugam, and I. Dinstein, 1973. Textural features for image classification: Institute of Electrical and Electronics Engineers Transactions on Systems, Man, and Cybernetics, SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69.

La Marca, K., 2020. Seismic Attribute Optimization with Unsupervised Machine Learning Techniques for Deepwater Seismic Facies Interpretation: Users vs Machines. Msc thesis. The University of Oklahoma.

La Marca, K., Bedle. H., 2022. Deepwater seismic facies and architectural element interpretation aided with unsupervised machine learning techniques: Taranaki basin, New Zealand. Marine and Petroleum Geology 136 – 105427.

Lima, B. E., Tedeschi, L., Pestilho, A. L., Santos, R., Vazquez, J., Guzzo, J., De Ros, L. 2020. Deep-burial hydrothermal alteration of the Pre-Salt carbonate reservoirs from northern Campos Basin, offshore Brazil: Evidence from petrography, fluid inclusions, Sr, C and O isotopes. Marine and Petroleum Geology. 104143. 10.1016/j.marpetgeo.2019.104143.

Lloyd, S. P.,1982. Least squares quantization in PCM. IEEE. Transactions on Information Theory, 28, 129–137.

Maas, M.V.R., Takenaka, G.O, Barreto, A.C.R., Sobreira, M.C.A., Pintas, E.M., Guerreiro, J.P.B.C., Chagas, A.A.P.,2019. Predição sísmica de litologias permoporosas na zona BVE-100 (Fm. Barra Velha) usando inversão Rp-Rs de Fatti em áreas exploratórias. Estudo de caso: Prospecto Araucária, Bloco Uirapuru, Pré-sal da Bacia de Santos. I Congresso Petrobras de Geociências e Geoengenharia. PETROBRAS. Rio de Janeiro.

Matos, M.C., Marfurt, K. J., Johann, P.R.S., 2010. Seismic Interpretation of Self-Organizing Maps using 2D color displays, 2010: Brazilian Journal of Geophysics 28(4): 631-642.

Matos, M.C., Yenugu, M., Angelo ,S.M., and Marfurt, K.J., 2011. Integrated seismic texture segmentation and cluster analysis applied to channel delineation and chert reservoir characterization," GEOPHYSICS 76: P11-P21. https://library.seg.org/doi/10.1190/geo2010-0150.1

Moreira, J.L.P., Madeira, C.V., Gil, J.A., Machado, M.A.P. 2007. Bacia de Santos. Rio de Janeiro. Boletim de Geociências da Petrobras, v. 15, n. 2, 531-549.

Penna, R., Lupinacci, W. 2020. Decameter-Scale Flow-Unit Classification in Brazilian Presalt Carbonates. SPE Reservoir Evaluation & Engineering. 23. 1-20. 10.2118/201235-PA.

Poros, Z., Jagniecki, E., Luczaj, J.,Kenter, J., Gál, B., Correa, T., Diniz Ferreira, E. L., Heumann, M., Elifritz, A., Johnston, M., Mcfadden, K., Matt, V. 2017. Origin of Silica in Pre-Salt Carbonates, Kwanza Basin, Angola. 10.13140/RG.2.2.31863.42405.

Oliveira, L.C., Rancan, C.C., 2019. Sill emplacement mechanisms and their relationship with the Pre- Salt stratigraphic framework of the Libra Area (Santos Basin, Brazil). LASI VI-The physical geology of subvolcanic systems: laccoliths, sills and dykes. Malargüe, Argentina.

Sartorato, A.C.L, Mizuno, T.A., Paegle, J.S., Pozzi, R.P.S, Zanatta, A.S., Rosseto, J.A., Oliveira, L.C., Penna, R.M., 2018. Modelagem geológica do reservatório do Campo de Mero. XI Seminário de Interpretação Exploratória. PETROBRAS. Rio de Janeiro.

Stica, J. M.; Zalan, P. V.; Ferrari, A. L., 2014. The evolution of rifting on the volcanic margin of the Pelotas Basin and the contextualization of the Paraná-Etendeka LIP in the separation of Gondwana in the South Atlantic. *Marine and Petroleum Geology*, [Amsterdam], v. 50, p. 21.

Taner, M. T., Koehler, F., and Sheriff, R. E. (1979), "Complex seismic trace analysis," GEOPHYSICS 44: 1041-1063.

Tritlla, J. & Loma, R., Sanchez Perez-Cejuela, V., Cerdà, M., Wloszczowski, D., Perona, R., Algíbez, J., Caja, M. A., Levresse, G., Bonillo, P., Tiwary, D. ,2013. Volcanism, alkalinity and hydrothermalism in offshore Brazil: unraveling the Pão de Açúcar reservoir. Repsol E&P Conference Madrid, Spain.

Zalan, P. V. ,2014. Comparisons between the Deep Crustal Structure of Magma-Poor and Volcanic Passive Margins. In: Atlantic Conjugate Margins Conference, 4. St. John's, Canada. *Extended Abstracts…* St. John's, Canada: [S.n.], 2014. p.75–78.

Zhang, K., Marfurt, K. J., and Guo, X., 2008. Volumetric application of skewed spectra: 77th Annual International Meeting of the SEG, Expanded Abstracts, 919-923.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015. A comparison of classification techniques for seismic facies recognition: Interpretation, 3, no. 4, SAE29–SAE58, doi: 10.1190/INT-2015-0044.1.

Chapter 3- Unraveling the hidden relationships between seismic multi-attributes, well dynamic data, and Brazilian pre-salt carbonate reservoir productivity: a shallow versus deep machine learning approach.

*This study is published in the SEG Interpretation journal as

Maas, M. V. R., Bedle, H., Ballinas, M.R., Matos, M. C. 2024. Unraveling the hidden relationships between seismic multi-attributes, well-dynamic data, and Brazilian pre-salt carbonate reservoirs productivity: A shallow versus deep machine-learning approach. Interpretation 12(4):1-61. https://doi.org/10.1190/int-2023-0113.1

Abstract

During the initial phases of an E&P project, the most reliable data on the reservoir's deliverability are acquired via drill stem tests (DST), which provide productivity per flow units, whenever production logging tool (PLT) data are available. However, DSTs are restricted to a few kilometers, whereas seismic data cover large areas. The integration of these data has been challenging, particularly due to the difference in scale between them. So, a new workflow to determine the relationship between post-stack seismic attributes and reservoir productivity using classic supervised (shallow) and deep learning regression algorithms was proposed. The DST parameters were predicted over the entire seismic cube, which can be extremely valuable for the decision-making process. The dataset is from the Brazilian deep water pre-salt carbonate reservoirs of the Mero Field, which is a well explored area with a plethora of test and production data. It is adjacent to an underexplored area (Central Libra appraisal plan), which is covered by the same seismic survey. Thus, any relationships between seismic attributes and well productivity data observed at the Mero field are extrapolated to the adjacent underexplored area. Ten seismic attributes and DST data from ten wells of Mero Field were used to train shallow and deep learning supervised regression algorithms for the prediction of flow capacity and productivity index seismic cubes. Twenty development wells (blind tests) were employed for the assessment of our predictive models. The highest percentage of correct predictions at the blind test wells (85%) was obtained with random forest regression using six attributes derived from a spectrally balanced full-stack volume, neither AVO nor inversion data were needed. Deep learning provided lower performance (75%) at a higher computational cost. It demonstrated a new reservoir de-risking tool that can be used for project optimization in areas covered by the same seismic survey.

Introduction

Before the commercial production of an oil field, drill stem tests (DST) are usually the most reliable reservoir productivity data. In a very simplistic view, a DST consists of production (or injection) of fluids from the perforated interval and measurements of rates and pressure variations. Its interpretation leads to two key parameters for reservoir characterization: total flow capacity (FC – also known as *kh*), and productivity index (PI). When a production logging tool (PLT) is also available, it is possible to observe the contributions of the layers within intervals, also known as flow units, for the total result (Bourdet, 2002). Depending on the reservoir nature, flow units identified through PLT have thicknesses compatible with the seismic resolution. This observation leads us to the challenge of determining the relationship between DST parameters and seismic multi-attributes to predict reservoir productivity as early as possible in the E&P projects timeline with an aim their speed up and optimization.

Some researchers have been working with reservoir prediction using well data (well logs and cores) and seismic inversion attributes (acoustic and elastic impedance) to primarily predict porosity and permeability, without using dynamic reservoir data from DSTs (Mitra and Khara, 2001; Taboada et al., 2001; Raeesi et al., 2012; Alfarraj and Alregib, 2018; Abdulaziz et al., 2019; Anifowose et al., 2019; Hosseini et al., 2019). In the Mero Field (Brazilian Deep Water – Pre-Salt), our study area, Penna and Lupinacci (2020) used seismic inversion attributes, well log, and core data to predict decameter scale flow units. These procedures are effective for field development optimization but cannot help the exploration teams that lack reliable seismic inversion data because of the low number (if not zero) of wells for an accurate wavelet estimation. In this work, we focus on understanding the relationships between easy-to-obtain seismic attributes and carbonate reservoir productivity at Mero Field. It is a continuation of the work of Maas et al., (2023), which qualitatively detected sweet spot reservoirs through unsupervised clustering of seismic multi-attributes. This time, we explored three classical (shallow) supervised machine learning (ML) algorithms (random forests-RF, support vector machines-SVM, and K-nearest neighbors - KNN), and deep neural network (multi-layer perceptron-MLP) for prediction of both FC and PI by training them with DST/PLT data of ten wells and ten different seismic attributes (instantaneous, geometric and AVO). The predictions were made over the entire seismic cube covering the Mero Field (in commercial production since April 2022 but producing since 2017) and Central Libra Appraisal Plan Area (CLAP), which currently is still in the exploratory stage. We used DST data from twenty development wells, as blind tests for our predictive model’s assessment as they were not used for any training.

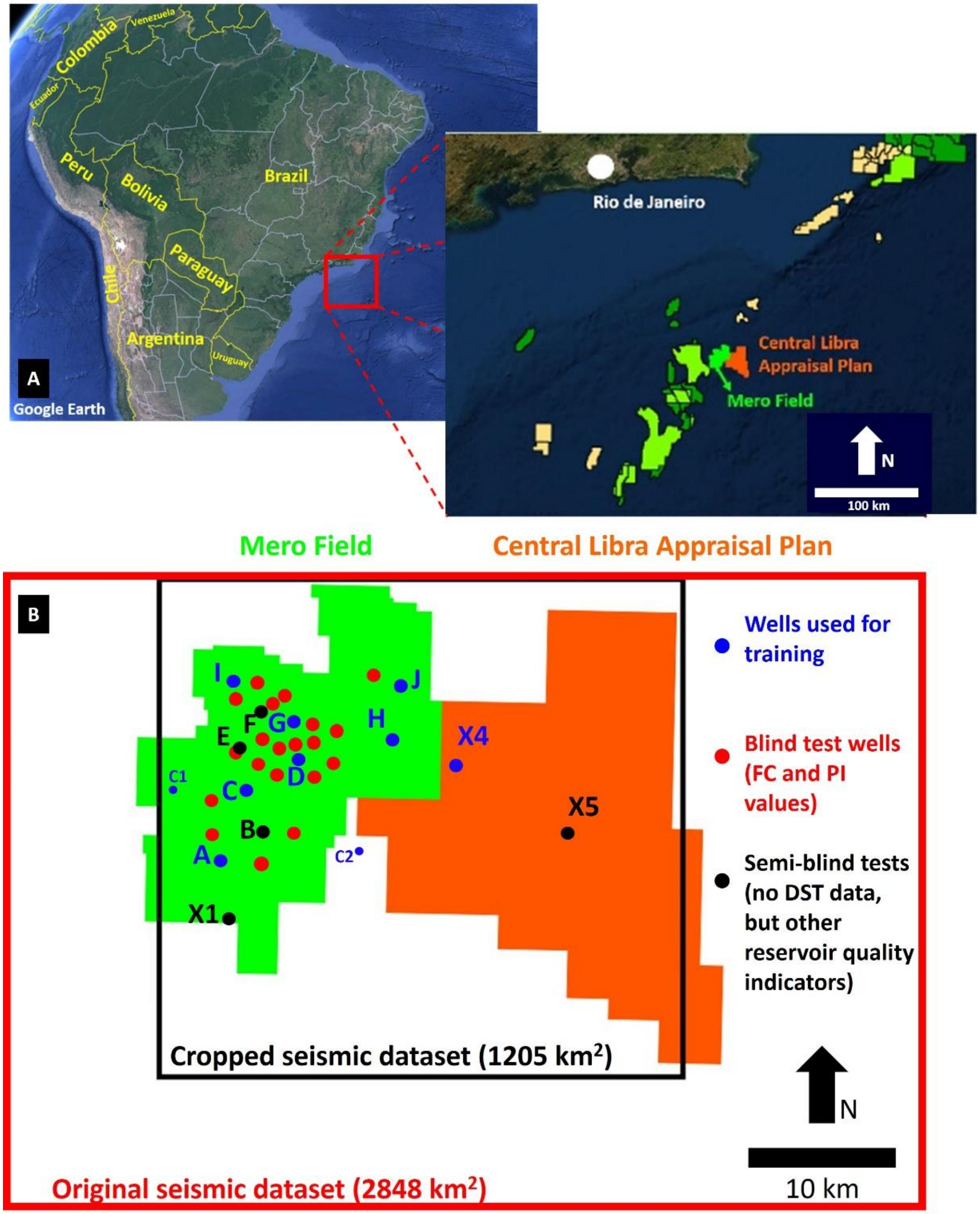

Figure 3.1. A) Location map of the Mero Field and Central Libra Appraisal Plan (CLAP) area; B) Wells and seismic dataset used in this study. Wells A to H have drill stem test (DST) data. Wells X1 to X5 were drilled for appraisal purposes. Wells A and C were selected as an injector-producer pair to support an anticipated production system test (Modified from Maas et al., 2023).

Area location and geological background

The Mero Field (Santos Basin deep waters, SE Brazil, Figure 3.1A) was discovered in 2010 and, since 2013 the operations have been conducted by a joint venture composed of Petrobras (operator), Shell Brasil, TotalEnergies, CNPC, CNOOC, and PPSA (Brazilian Government Representative), commercial production started in 2022. It is adjacent to the CLAP area, which has only two exploratory wells drilled (one discovery and one dry well). As both areas are covered by the same seismic survey, any machine learning pattern detected at Mero Field can be immediately extrapolated to the underexplored adjacent area (Figure 3.1B) (Maas et al., 2023).

The Santos Basin evolved from a continental rift in the Early Cretaceous (Neocomian/Barremian) to a passive margin in the Late Cretaceous and Tertiary (Moreira et al., 2007) (Figures 3.2A and 3.2B). Further details on Santos Basin tectonostratigraphic evolution can be found in Moreira et al., (2007); Zalan, et al., (2014), and Stica (2014). The main reservoirs of the Mero field are Late Aptian post-rift lacustrine carbonates of the Barra Velha Formation (BVE), which are the focus of this analysis. The most important producing facies are build ups and mounds composed mainly of shrubby stromatolites over basement horsts and adjacent to deep faults (Figures 3.2A and 3.2C). Secondarily, there are producing lithologies related to platform facies composed of grainstones and rudstones (reworked from adjacent build ups and mounds) intercalated with algal laminites and spherulitites. In the depositional lows, there are distal facies (non-reservoirs) composed mainly of mudstones and magnesium rich clays (Sartorato et al., 2018) (Figure 3.2C). In the Mero Field, the BVE lithologies are intruded by Santonian age alkaline igneous bodies with negligible thermal effect on the reservoir (thin metamorphic aureole), and the intrusive sills themselves act as flow barriers within the reservoir (Oliveira and Rancan, 2019). A pervasive hydrothermal alteration of the reservoir is characterized by moderate to strong silicification mainly within the build ups and mounds close to deep faults (Figure 3.2A). It is a regional process that is well documented in literature by Lima et al., (2020); Tritla et al., (2013); and Poros et al., (2017). While these studies advocate a regional hydrothermal system, other researchers (Oliveira and Rancan, 2018; Kangxu et al.,2019, Oliveira and Rancan, 2019, Oliveira et al., 2024) associated the hydrothermal vents observed at Mero Field seismic data with the Santonian igneous intrusions. The exact timing of the hydrothermal event is still under investigation. At most of Brazilian pre-salt fields, this silica-rich hydrothermal alteration has a strong positive effect on the reservoir productivity. The most productive wells were drilled in strongly silicified and karstified carbonate build ups and mounds (Sartorato et al., 2020). However, some severe drilling issues occur within those highly productive reservoirs, like severe circulation losses due to the cave system and very low rates of penetration due to the silicified cave walls (Rubim et al., 2022; Gonçalves et al., 2022, Petrobras internal reports). In some cutting samples descriptions, both microcrystalline and cryptocrystalline quartz are reported, indicating that, in some places, the giga pores are metric to decametric cavities with quartz crystal walls. Thus, pre-drilling prediction of these highly productive and hard-to-drill reservoirs is important for both field operations and project profitability optimization.

Dataset

We analyzed 3D seismic data and dynamic reservoir parameters from twenty-seven wells drilled at Mero Field and CLAP area (Figure 3.1B). It is the same dataset used by Maas et al., (2023) with the addition of fourteen more development wells that were used for supervised machine learning predictive models' assessment.

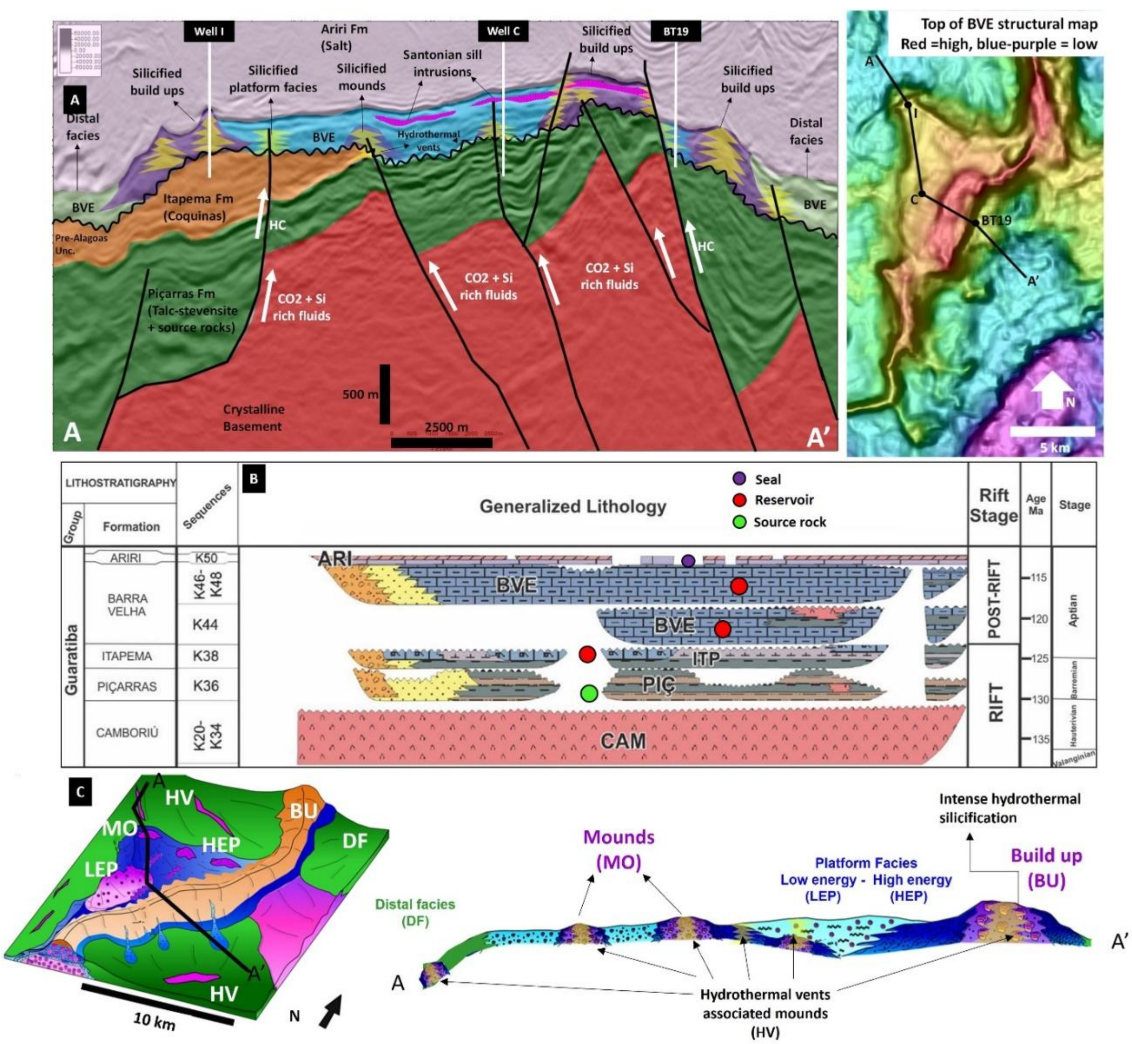

Figure 3.2. A) Structural section interpreted from original seismic amplitude volume (RTM depth migrated). The section trajectory is plotted over the top of the BVE reservoir surface. Section A-A' passes through wells I (training well, hyper producer), C (training well, good producer), and BT19 (blind test well, highly injective). Well C penetrated highly productive carbonate platform facies (mainly grainstones made by build ups reworking and laminites) close to a hydrothermal vent associated with deep faults. Injection well BT19 was drilled at fault footwall, is highly injective, and the

main reservoir is made by gravitational deposits from build up reworking. White arrows indicate paths for hydrocarbons (HC), carbon dioxide (CO_2), and silica (Si) rich fluids through reactivated faults that acted as hydrothermal vents. B) Chronostratigraphic chart of Santos Basin: BVE reservoir is composed of sequences K44, K46, and K48 (modified from Moreira et al., 2007). C) Our depositional model of the BVE reservoir at Mero Field (modified from Sartorato et al., 2018).

Seismic data

We derived classical seismic attributes described by Chopra and Marfurt (2007) from depth migrated (Reverse Time Migration - RTM) full and angle stacks that cover both the Mero Field and CLAP area (Figure 3.1B). In addition to these classical attributes, from angle stacks, we also used as input Fatti's AVO attribute, which mimics a seismic bandwidth lambda-rho (Fatti et al., 1994). This seismic input suite (full stack and angle stack derived attributes) is the most common in areas at exploratory stage, where reliable seismic inversion data is absent (Maas et al., 2023).

Well data

We used data from eight exploration and appraisal wells for the training dataset and data from twenty development wells as blind tests for prediction assessment (Figure 3.1B). Wells A to J have both electric logs, rock and fluid samples, and drill stem test data (DST). Wells X1, X4, and X5 have only electric logs and rock/fluid samples. Wells A (injector), C, H, and BT17 (blind test) have also extended DST data (44 months) and production history (anticipated production system). The main DST parameters are flow capacity (FC= transmissibility * fluid viscosity), productivity index (PI= flow rate/pressure variation), and production logging tool (PLT) flow units, which are individual or multiple permeable reservoir units that contribute to well productivity.

Method

The workflow is summarized in Figure 3.3 and will be detailed in the next subsections. Each of the seismic data volumes is 17 GB in size. For this reason, they were cropped around the area of interest (AOI) in the Mero Field and CLAP area (Figure 1B). After cropping, each seismic volume is reduced to 2 GB size, and all of them were analyzed and interpreted in the depth domain. The seismic grid is 25 x 25 x 5 m, other acquisition and processing parameters can be found in Maas et al., (2023).

Seismic data conditioning, and attribute extraction

The analysis started with a basic seismic data interpretation workflow, which comprises seismic-well tie, top and bottom of BVE reservoir horizon picking, and finally, attribute computation. From the full stack volume, we derived nine seismic attributes: Hilbert transform of amplitude (HTAM), sweetness (SWEET), energy ratio similarity (ERS), average frequency (AVF), gray level co-occurrence matrix energy- GLCMs (energy-GLEN, entropy-GLET, homogeneity-GLEH), maximum positive (K1), and maximum negative (K2) curvatures. From the angle stacks, we calculated the Hilbert transform of Fatti's lambda-rho (HTLR- an AVO attribute). Detailed descriptions of each attribute can be found at Chopra and Marfurt, (2007). The reasoning for the selection of these attributes can be found at Maas et al., (2023). They found that instantaneous (HTAM, AVF, SWEET), geometric (K1, K2, ERS), pixel-based texture (GLCMs), and AVO (HTLR) attributes are geologically meaningful for facies prediction at BVE reservoirs using unsupervised clustering algorithms. Figure 3.4 shows 3D views of each one of the ten input seismic attributes at the top of the BVE reservoir.

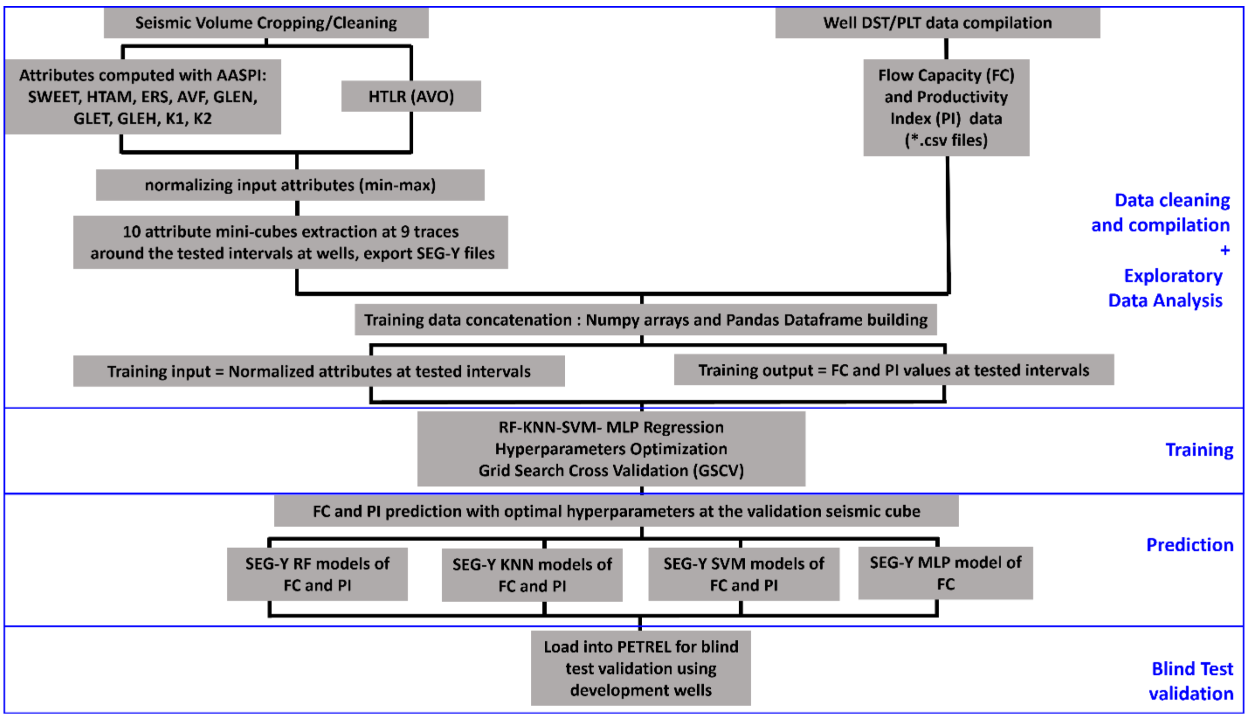

Figure 3.3. Workflow used in this study. Acronyms: DST (drill stem test), PLT(production logging tool), AASPI (attribute assisted seismic interpretation), SWEET (sweetness), HTAM (Hilbert transform of amplitude), ERS (energy ratio similarity), AVF (average frequency), GLEN (gray level co-occurrence matrix energy), GLET (gray level co-occurrence matrix entropy), GLEH (gray level co-occurrence matrix homogeneity), K1 (maximum positive curvature), K2 (maximum negative curvature, HTLR (Hilbert transform of Fatti's lambda-rho), AVO (amplitude versus offset), FC (flow capacity), PI (productivity index), RF (random forest), KNN (K-nearest neighbors), SVM (support vector machine), GSCV (grid search cross validation), MLP (multi-layer perceptron).

Well dynamic data compilation and training dataset structuration

The Mero field wells with DST and/or PLT data at BVE formation were separated into three groups (Figure 3.5): i) mild producers, mostly related to lower energy platform facies: their typical FC values fall between 1,000 and 10,000 mD.m. The PI values are less than 100 m^3/d.kgf.cm^2. The average oil production at BVE is 8,000 barrels of oil per day (bpd). ii) good producers related to high energy platform facies and low energy facies close to hydrothermal vents or mounds: FC values between 10,000 and 85,000 mD.m, PI values between 100 and 500 m^3/d.kgf.cm^2, and average daily oil production at BVE of 24,000 bpd. iii) hyper producers related to karstified and silicified mounds and build ups. Their FC values explode to hundreds of thousands of mD.m and the PI values also explode to more than 1000 m^3/d.kgf.cm^2. Wells like these have the highest oil flow rates observed at BVE formation (48,000 bpd). Note that as FC and PI are strongly linearly correlated (R^2 = 94.7%), both can be targeted as predicted outputs (Figure 3.5). It is important to mention that the FC and PI values are not restricted to depositional facies. Despite the hyper flow units being most common within the silicified build ups and mounds, they also occur within the platform and distal facies close to hydrothermal vents. One good example is Well C, it drilled silicified reworked grainstones close to deep faults that acted as pathways to hydrothermal silica-rich fluids (Figure 3.2A). As a result, its FC and PI values (50,000 mD.m and 370 m^3/d.kgf.$cm^{2,}$ respectively) are significantly above the average of the non-altered platform facies (FC < 10,000 mD.m and PI < 100 m^3/d.kgf.cm^2).

Among the two first groups of wells, the platform facies ones, BVE carbonates are mainly composed of interbedded layers of in situ spherulitites, microbial mats, and grainstones (Figures 3.2A and 3.2C). Due to the layered nature of the reservoir, the horizontal permeability is higher than the vertical one. Consequently, wells that drilled such reservoirs often show vertically isolated flow units that can only be indicated through production logging tools (PLT). At the hyper productive reservoirs from the silicified build ups and mounds, the dissolution caused by at least two karstification events (one meteoric and one hydrothermal) yielded a metric to decametric scale highly interconnected cave system which makes both vertical and horizontal permeability equally extremely high. Therefore, at the very porous and permeable build ups and mounds, no PLT is required once the whole tested interval is contributing to oil flow. Souza et al., (2021) describe the corrosive effect of silica-rich fluids when they percolate layered carbonate reservoirs.

For building the training dataset, we compiled FC and PI information from wells A, C, D, G, H, I, and J. We note that Well X4 found no reservoir at BVE, just intercalation of distal mudstones with Santonian intrusive igneous bodies. To avoid biased information in the training dataset due to the existence of only productive wells DST data, we assume that both FC and PI values are zero at very low porosity mudstones and igneous intrusions. For this reason, we included Well X4 and pseudo-wells C1 and C2 as non-producers to counterbalance the potential biasing in the training dataset (Figure 3.6A). In the Mero field, most of the flow units identified through DST are compatible with the seismic resolution of the area (25 - 40m) (Figure 7). For this reason, we can use seismic attributes for FC and PI prediction for the whole seismic cube, within BVE interval. At each well with DST/PLT data, at the tested interval depths, we extracted seismic mini-cubes (9 traces around the well) of each one of the ten input seismic attributes. Therefore, for each one of the ten wells, at the tested intervals of the BVE reservoir, we have mini-cubes (prisms) of ten seismic attributes, flow capacity, and productivity index, totalizing 4968 data points in the training dataset

(Figures 3.6B, 3.6C and 3.6D). As the seismic sample interval in the z direction is 5 m, and the seismic grid is 25x25 m in the x and y directions, the number of training voxels per well ranges from 270 at the thinnest perforated zone (145 m at Well G) to 657 voxels at Well C (360 m thick perforated zone). These seismic mini-cubes were converted into SEG-Y files, then into NumPy arrays, and, finally into a Pandas data frame which is composed of ten rows, each one related to the input seismic attribute (independent variable or feature) and another row with the predicted output (FC or PI values from DST/PLT) (Table 3.1).

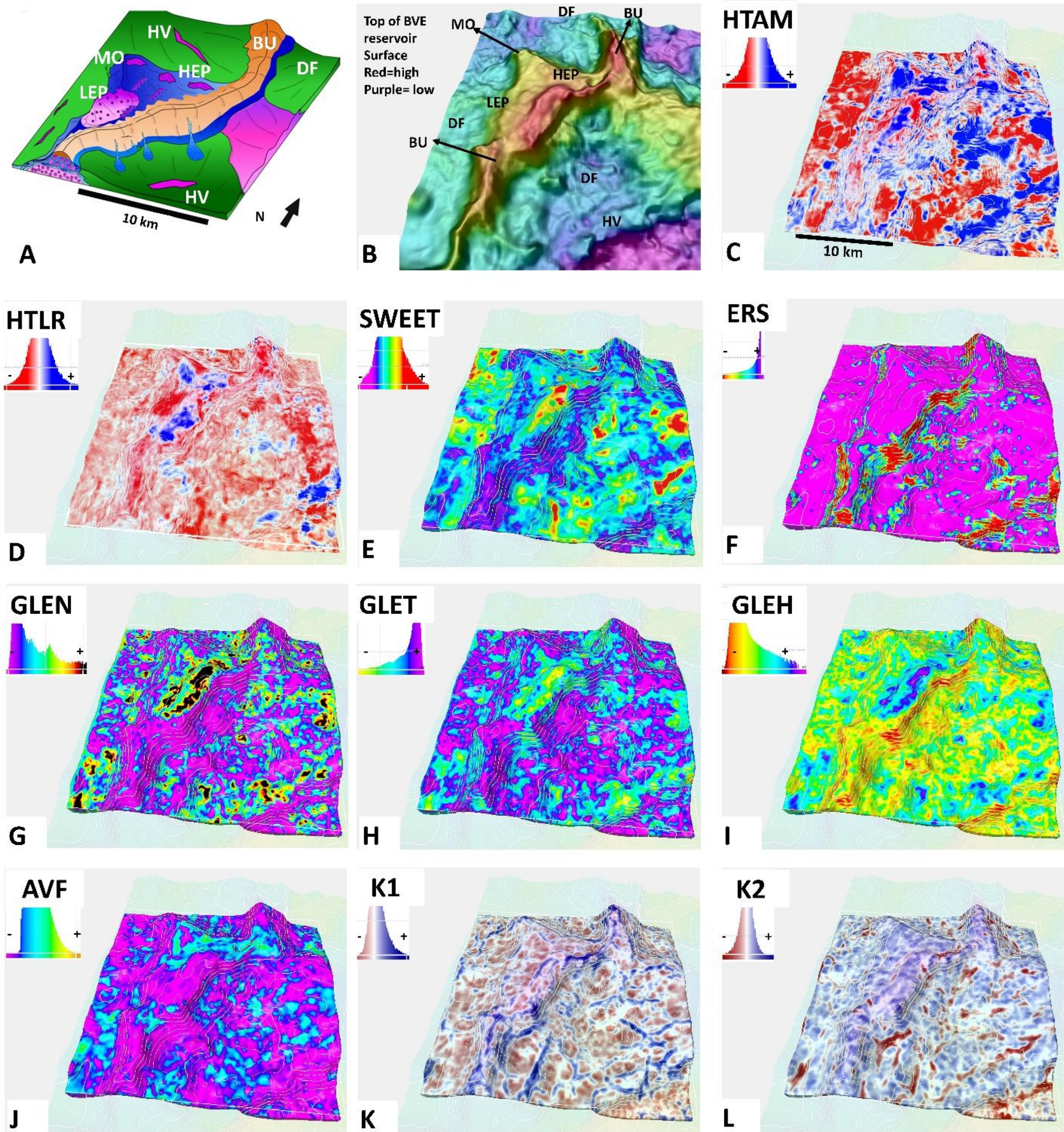

Figure 3.4. 3D view of the seismic attributes used in this study at the top of Barra Velha reservoir surface. A) Depositional model of BVE reservoir at Mero Field (modified from Sartorato et al., 2018); B) Top of BVE reservoir surface; Seismic attributes: C) Hilbert transform of amplitude; D) Hilbert transform of Fatti's Lambda Rho; E) Sweetness; F) Energy ratio similarity; G) GLCM energy; H) GLCM entropy; I) GLCM homogeneity; J) Average frequency; K) Maximum positive curvature- K1; L) Maximum negative curvature-K2. Geologic features acronyms: BU (Build ups), MO (Mounds), HEP (high energy platform facies), LEP (low energy platform facies), DF (distal facies), HV (hydrothermal vents), (Modified from Maas et al., 2023).

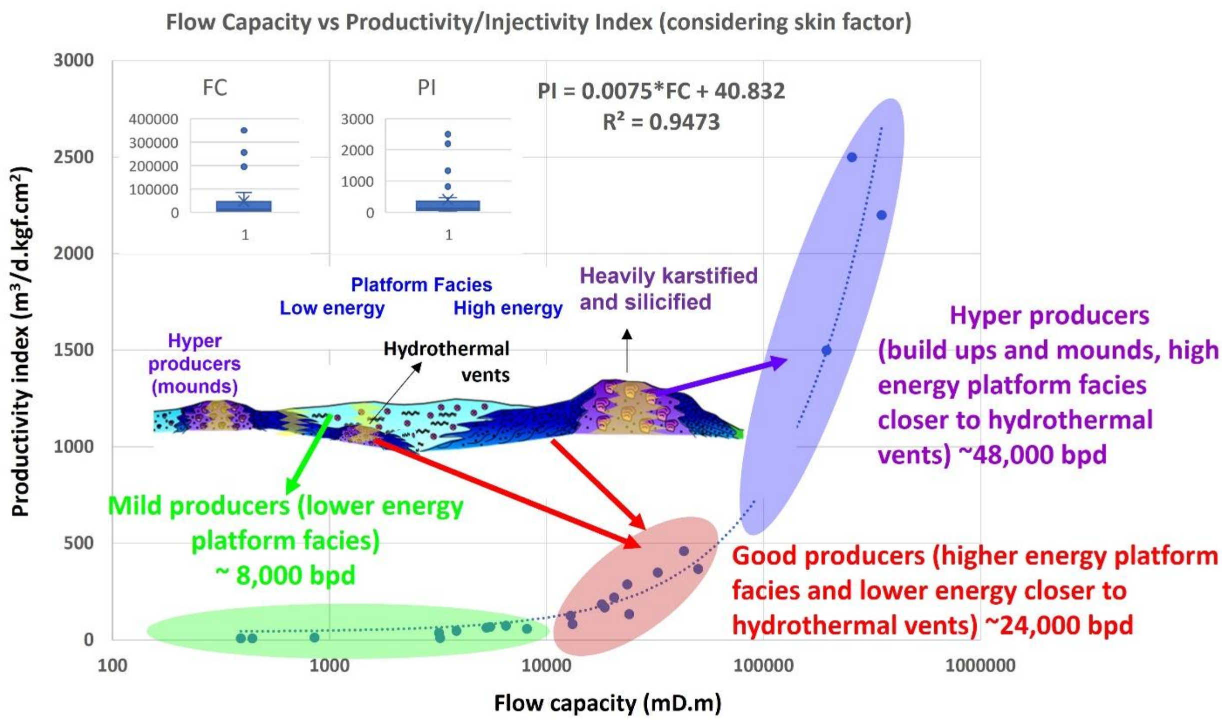

Figure 3.5. DST data compilation at BVE formation at Mero Field. This cross plot shows the linear relationship between flow capacity (FC- x axis, logarithmic scale) and productivity index (PI-y axis) obtained through DST. At the left top corner, we have the box plots of FC and PI respectively. Most of the DST data were acquired at mild (8,000 bpd oil production) and good producers (24,000 bpd) from low and high energy platform facies respectively. When closer to hydrothermal vents, these platform reservoirs have higher FC and PI values. The hyper productive wells from silicified build ups and mounds represent anomalies related to very high transmissibility reservoirs characterized by karstified and fractured stromatolitic accumulates. At the Mero field, these sweet spot reservoirs can produce up to 48,000 bpd (BVE only). Note that both FC and PI are not totally related to depositional facies, the hydrothermal events altered all facies and, as a result, enhanced pre-existing transmissibility, mainly at the build ups and mounds.

Exploratory data analysis

First, we applied min-max (0-1) normalization to each one of the ten input attributes. With Pandas library functions, we calculated the main statistical parameters such as mean, correlation, and covariance. Figure 3.8A shows a heatmap of the correlation between each one of the ten input attributes and flow capacity. It can be observed that the correlation is very low (close to zero) for all the attributes, this indicates that with one attribute at a time, it is not possible to predict flow capacity. When comparing the attributes, the correlation between them is also low. The exceptions are the GLCM attributes which have high correlation (0.6 – 0.9). However, where the correlation is lower than 0.75, they still complement each other and, therefore, are not redundant. This makes them useful for contributing individually to machine learning based predictions of FC and PI.

In Figure 3.8B, we present a pair plot (Seaborn library) of some of the normalized input attributes, the hue is the FC value. Note that for almost all the attributes, there is no linear separability. The only exception again is related to the GLCM attributes given their higher correlation. Figure 3.9 contains the histograms of each one of the ten input attributes and the flow capacity values were grouped into the three distinct categories mentioned before (0-10,000 mD.m, mild producers; 10,000-85,000 mD.m, good producers; >85,000 mD.m, hyper producers). We can observe, again, that with one attribute at a time, it is impossible to predict flow capacity, as there is significant overlapping between the three categories related to seismic ambiguity. For human beings, it is impossible to confidently predict FC or PI with each one of these attributes separately due to three big issues: data is not linearly separable, data is overlapping and there are a lot of independent variables (ten attributes) configuring an 11D feature space (hyperdimensional). Only machine learning algorithms can cope with such a multidimensional dataset and detect subtle patterns that are not perceptible to the most experienced human seismic interpreters.

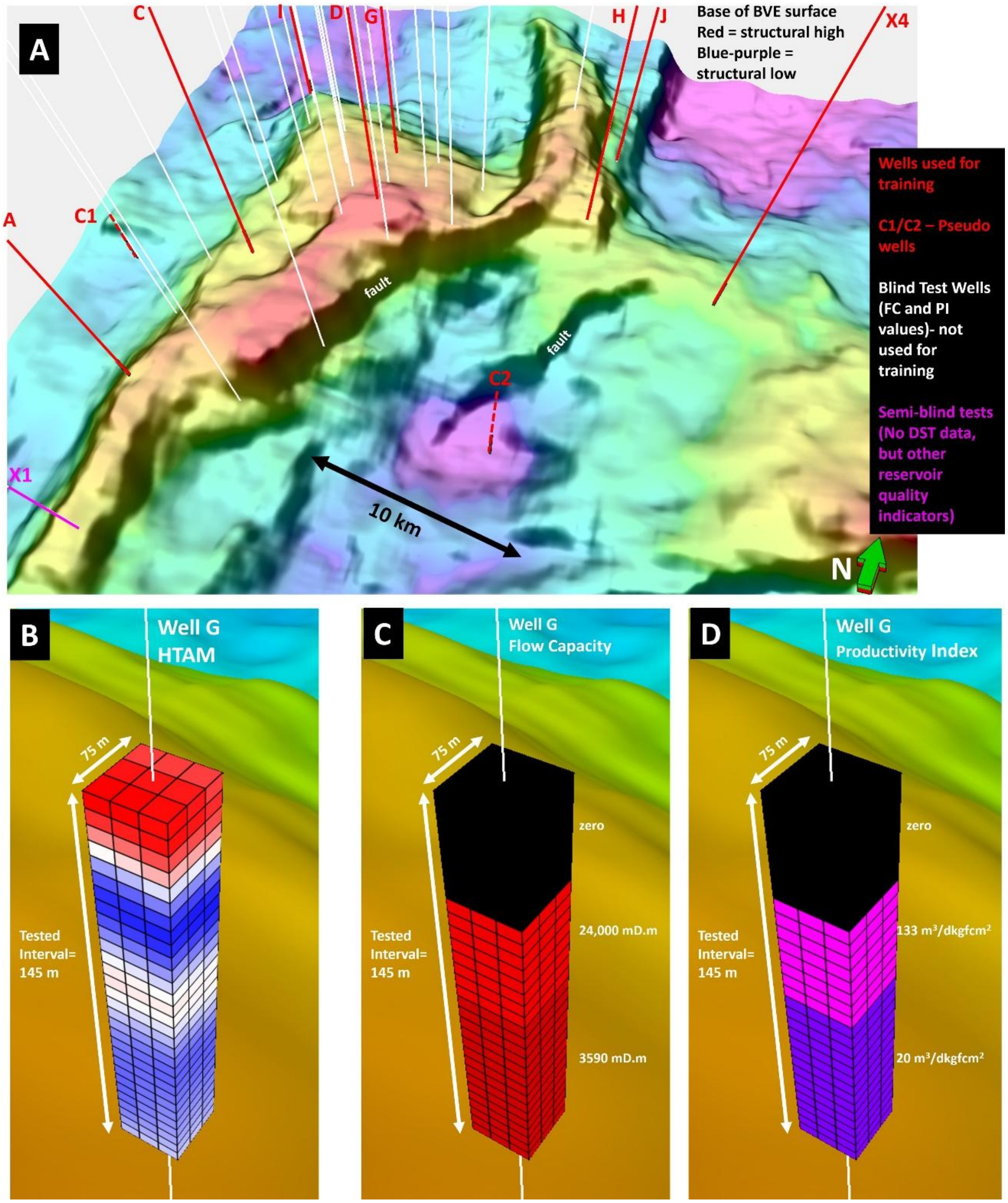

Figure 3.6. A) 3D view of the wells used in this study. The surface is from the base of the BVE reservoir with faults. In the Mero Field, the BVE reservoir is deformed by extensional reactivation of rift faults but with relative preservation of the original platform geometry. Wells in red are mostly appraisals and were used for training the supervised machine learning algorithms used in this study. Wells C1 and C2 are pseudo-well locations where BVE has presumed non-reservoir facies. Wells in white are development wells that have both FC and PI data, they were not used for training and, thus, consist of blind tests for quantitative assessment of our prediction models. Wells in pink do not have DST data but have other reservoir quality indicators (mainly mini-DST) and, thus, also can be used as a semi-blind test for qualitative assessment of our predictive models. From B to D, we have 3D views of one example of how the training dataset is built. At the tested interval, we extract seismic mini-cubes (prisms) around the well (nine traces) (Modified from Maas et al.,2023); B) seismic mini-cube of HTAM attribute along the tested interval (145 m) at the BVE reservoir, the sample interval in the z-direction is 5 m.

At the perforated zone, we have 30 voxels in the z direction and 9 voxels (3x3 voxels=75x75 m) in the x and y directions, totaling 270 voxels or data points around the perforated interval; C) The same mini-cube has the FC values for each flow unit identified via PLT. D) The mini cube has the values of PI for the same flow units. The surface is from the base of the BVE reservoir. We did the same for the other nine training wells, totaling 4968 training data points.

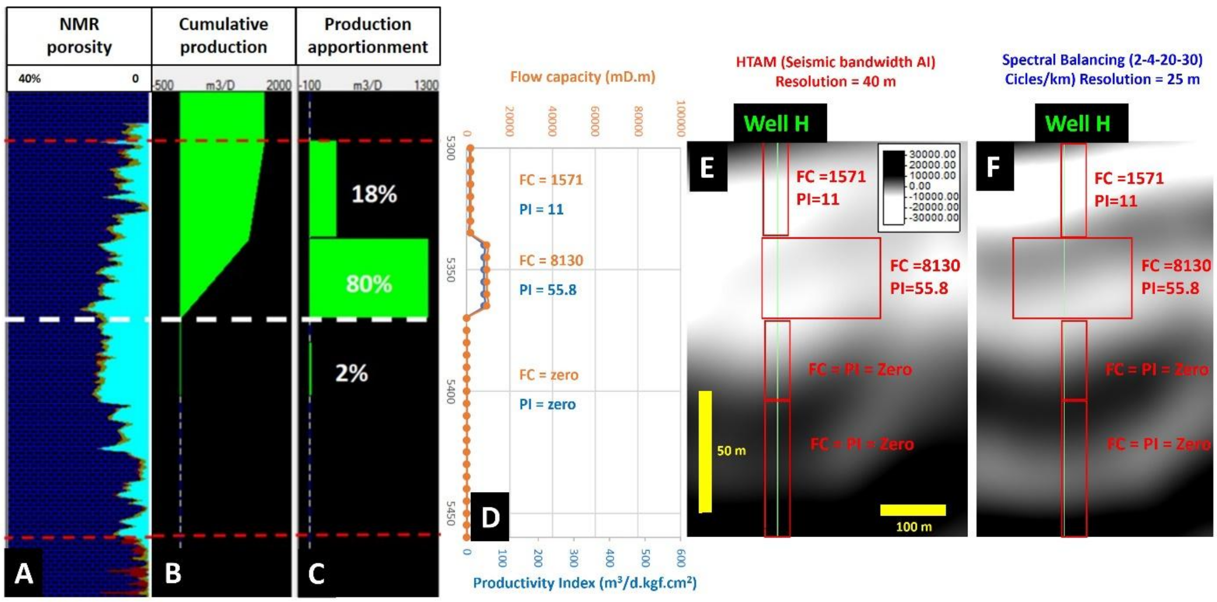

Figure 3.7. PLT flow units and seismic attributes scale matching. A) Nuclear magnetic resonance (NMR) porosity log at DST interval at Well H; B) Cumulative production curve obtained with production logging tool (PLT) at the same interval; C) Production apportionment obtained with the derivative of the previous curve: the steepest gradients at curve B are related to the main flow units that contribute to the cumulative production. Note that there are some zero flow intervals, where the production at curve B is flat; D) We modulated curve C by both DST values of FC and PI and used them for training dataset building; E) Hilbert transform of input amplitude (HTAM) seismic attribute at well H location: this attribute mimics a seismic bandwidth acoustic impedance. We can observe that the flow units (red polygons) identified through PLT have thicknesses that are compatible with the seismic resolution (40 m); F) Same attribute as in E but with spectral balancing which increased the seismic resolution to 25 m. We also have evaluated the impact of spectrally balanced inputs on our prediction models.

	HTLR	SWEET	K1	K2	ERS	GLEN	GLEH	GLET	HTAM	AVF	FC
0	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.001
1	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.001
2	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.001
3	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.001
4	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.001
...	...	...	...	...	...	...	...	...	...	...	...
4963	0.383269	0.055985	0.347361	0.597309	0.835561	0.147224	0.328122	0.869574	0.557484	0.261863	0.000
4964	0.391628	0.047381	0.348552	0.598968	0.786645	0.139299	0.260121	0.888977	0.550354	0.241475	0.000
4965	0.403035	0.039802	0.349424	0.600760	0.728800	0.129791	0.268291	0.908395	0.543518	0.222922	0.000
4966	0.413588	0.033232	0.350203	0.602580	0.662938	0.149249	0.268974	0.904110	0.537592	0.205748	0.000
4967	0.388991	0.000000	0.212625	0.582456	0.000000	0.000000	0.000000	0.000000	0.514293	0.051375	0.000

4968 rows × 11 columns

Table 3.1–A view of the Pandas data frame of the training dataset. Here we have all data points information of normalized seismic attributes and flow capacity at the BVE reservoir tested intervals of the Mero Field.

Basic concepts of selected supervised machine learning algorithms

First, we selected three of the most used classical (shallow) supervised regression algorithms: random forest (RF), support vector machines (SVM), and K-nearest neighbors (KNN). These algorithms have been used for different types of multivariate regression and classification when the independent variables are not linearly separable and contain higher dimensional data (Fukami et al., 2020; Rodriguez Perez and Bajortah, 2022, Ballinas et al., 2023). Additionally, for deep versus shallow learning comparison purposes, we used multi-layer perceptron (MLP) deep neural network for regression. With each one of the four methods, we modeled the relationship between a dependent variable (the target variable, in our case, FC and PI) and multiple independent variables (features, in our case, seismic multi-attributes). The cost function used for all methods maximizes the R-squared (R^2) between the measured and predicted values of FC and PI. The high dimensionality and non-linearity of the dataset require this kind of approach to model complex relationships between variables. For algorithms coding in Python language, we used the open-source library Scikit-Learn.

Random forest

Random forest regression (Ho, 1995) is a supervised machine learning algorithm and an ensemble learning method that combines multiple decision trees to make predictions of the desired output. Each decision tree in the forest is built using a random subset of the training data and of the input features (bootstrapping). For regression tasks, the final prediction is done by averaging the predictions of each one of the decision trees (aggregation). This procedure is commonly named in the literature as *bagging* (bootstrapping + aggregation). This algorithm architecture is interesting because, at an individual decision tree, not all input features (seismic attributes in this study) are used for training. This yields an important output of RF which is features (attributes) importance. Once correct predictions (high R^2) were made with some attributes, other decision trees predicted the target property (FC and PI in this study) with very low R^2. Thus, the algorithm can rank the input attributes in terms of their importance for good predictions (Figure 3.10A).

Below are the main hyperparameters of random forest regression:

Number of estimators: It is the number of individual decision trees in the random forest. The higher the number of estimators, the higher the performance of the prediction, but it also can lead to high computational cost and/or overfitting, i.e., it works well in predicting within the training dataset, but it is not good in generalizing for other datasets.

Maximum depth: It is the deepest level in the tree composed of only leaf nodes (Figure 3.11A). Deeper trees have the power to capture complex relationships but are also prone to overfitting.

K-nearest neighbors

K-nearest neighbors (KNN) is a non-parametric algorithm used for both classification and regression tasks (Fix and Hodges, 1951). In regression tasks using KNN, the prediction of the output parameter for a given data point in the training dataset is made by computing the average of the values of its K nearest neighbors, where K is the main hyperparameter. A secondary hyperparameter is a distance metric. In our case, we used Euclidean distance for the nearest neighbor's calculation. Very low values of K lead to model overfitting, whereas very high values cannot capture all the variability in the training dataset (Figure 3.10B).

Support vector machines

Support vector machines (SVM) is a supervised learning algorithm that can also be used for regression tasks (Cortes and Vapnik, 1995). In a training dataset, the algorithm builds a hyperplane that contains the most distant points from given data points (support vectors) and at the same time minimizes the error of the predictions (Figure 3.10C).

The main hyperparameters of SVM regression are:

C: This is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the prediction error. Low values of C lead to a larger margin but may result in higher error, on the other hand, high values of C lead to lower errors and smaller margin.

Epsilon: Is a tolerance parameter that defines the margin of acceptable errors. Data points that are located within tolerance are not penalized in the loss function. High values of epsilon lead to a wider error margin (Figure 3.10 C).

Kernel: in SVM regression, kernel functions are employed to map the data points into an emulated higher-dimensional feature space. The main kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid kernels. As our seismic attributes in the training dataset are non-linearly separable, we used RBF (Figure 3.10D).

Gamma: It defines the influence of each training data point in the hyperplane. Low values of gamma lead to smoother hyperplanes, whereas higher values make them more complex and, thus, can lead to overfitting.

Multi-layer perceptron

Despite it being a relatively old concept as it was first proposed by Rosenblatt (1958) in his neurobiological studies, it was implemented almost 30 years later by Rumelhart et al., (1986) using the new concept of backpropagation for neural network optimization. Multi-layer perceptron (MLP) is a typical example of a feedforward, fully connected, multi-layer, artificial neural network. The main hyperparameters are the number of hidden layers, the number of neurons at each hidden layer, and the activation function for each hidden layer. For classification tasks, the output layer has multiple neurons according to the number of classes. For regression problems (as the one described herein), the output layer is made of a single neuron in which the predicted value of the target variable is stored. Figure 3.10E shows the architecture of the MLP used in this supervised regression study. Each hidden layer is fully connected by weights and biases that are used to fit the input features (in this case, seismic attributes) to the desired output, in this case, flow capacity (FC) from drill stem tests (DST). The weights and biases are optimized during backpropagation when the R^2 between actual and predicted FC is maximized.

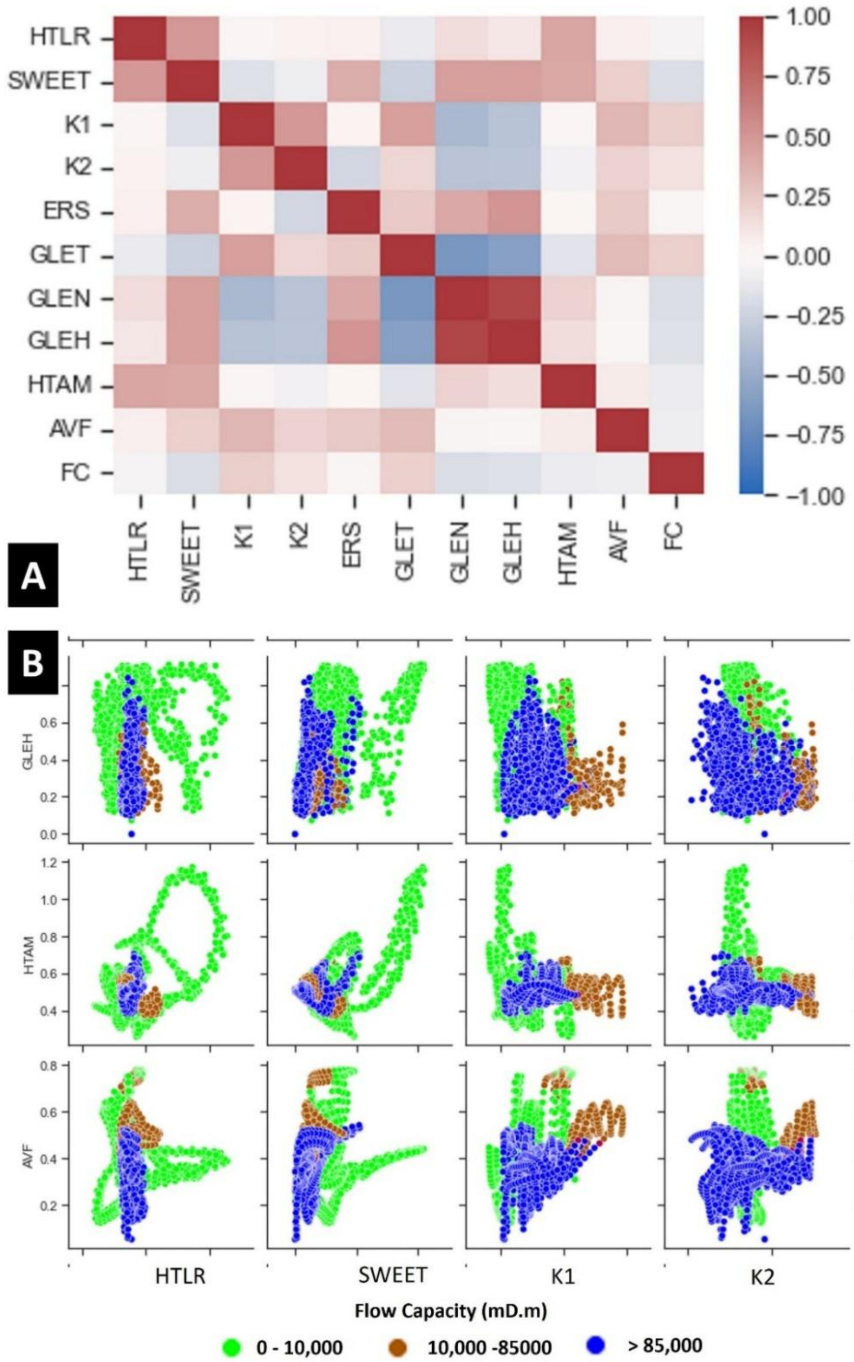

Figure 3.8. A) Heatmap plot of the ten input attributes and flow capacity at the tested intervals of BVE at the training dataset wells. Note that all the input features (attributes) have very low correlation among them. The exception is the GLCM ones, but even though they have correlation values around ± 60% (GLET), which makes them complimentary to each other. Another interesting aspect is the very low correlation between each of the input features with FC, which indicates that with just one attribute at a time, it is impossible to predict FC. B) Seaborn pair plot of some of the normalized input seismic attributes and the flow capacity values (hue) at the tested intervals of the training dataset wells. We can observe that most of the input features (except GLCM ones which are relatively highly correlated) are non-linearly separable. It is not possible to separate the hyper flow units (blue dots) from the other ones.

Data training, reservoir productivity prediction, and models validation

We started with random forest (RF) regression of FC using all ten attributes (HTAM, HTLR, ERS, AVF, SWEET, K1, K2, GLET, GLEN, GLEH). The reason for this algorithm choice is its intrinsic characteristic: as the training dataset is randomly redistributed among hundreds of decision trees with some input features (attributes) missing, it gives us an attribute importance rank. The random forest algorithm was trained using different combinations of hyperparameters through a 5-fold grid search cross-validation (GSCV). During 5-fold GSCV, the training dataset is split into five equal parts or folds, each fold is treated as a test set and the model is trained with the remaining four folds. This process is done five times by randomly picking different test subsets (shuffle). As the total number of training data points (voxels) is 4968, in each fold there are randomly picked 994 voxels used as tests, and 3974 voxels for training. As mentioned before, the maximum number of training voxels around a perforated interval is 657. Because of the randomness, it is not guaranteed that the test is performed with random voxels from different well locations, information leakage can occur and might lead to overfitting. The final fit of the model to training data is made with the hyperparameters which lead to the highest R^2 of FC and PI prediction during 5-fold GSCV (Brunton and Kutz, 2019; Fukami et al., 2020). After hyperparameters tunning, the prediction is made within the BVE reservoir at the entire seismic survey which covers both the Mero field and CLAP area. For investigation of the effect of dimensionality reduction, we also run RF regression of FC with 9, 8, and 6 input seismic attributes according to their importance. Next, with six attributes (AVF, K1, K2, ERS, HTAM, and GLET), we performed supervised regressions of both FC and PI using also SVM and KNN algorithms for comparison of their performance. For investigation of the effect of a higher seismic resolution input on model predictivity, we run RF regression of FC using the same six attributes derived from a spectrally balanced full stack volume. Finally, we run MLP regression (deep learning) of FC with ten attributes for cost/benefit comparison purposes with shallow learning models. So, a total of eleven supervised machine learning models were predicted in the studied area (Table 3.2).

The eleven supervised regression models were converted into SEG-Y volumes and loaded into a seismic interpretation program for prediction assessment using FC and PI values obtained through DST and PLT in twenty development wells (blind tests), which were not used for any training.

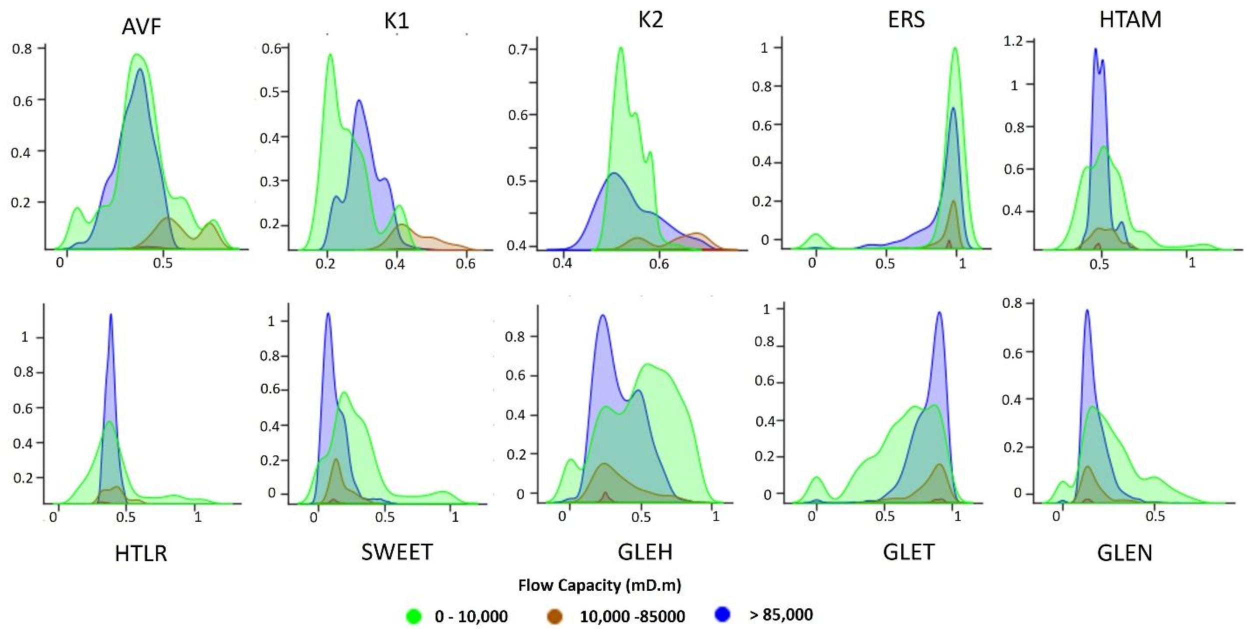

Figure 3.9. Histograms of the ten normalized input seismic attributes and flow capacity values at tested intervals of the training dataset wells. Here we can observe strong overlapping of FC values for each input attribute. Again, with one attribute at a time, we cannot predict FC or PI.

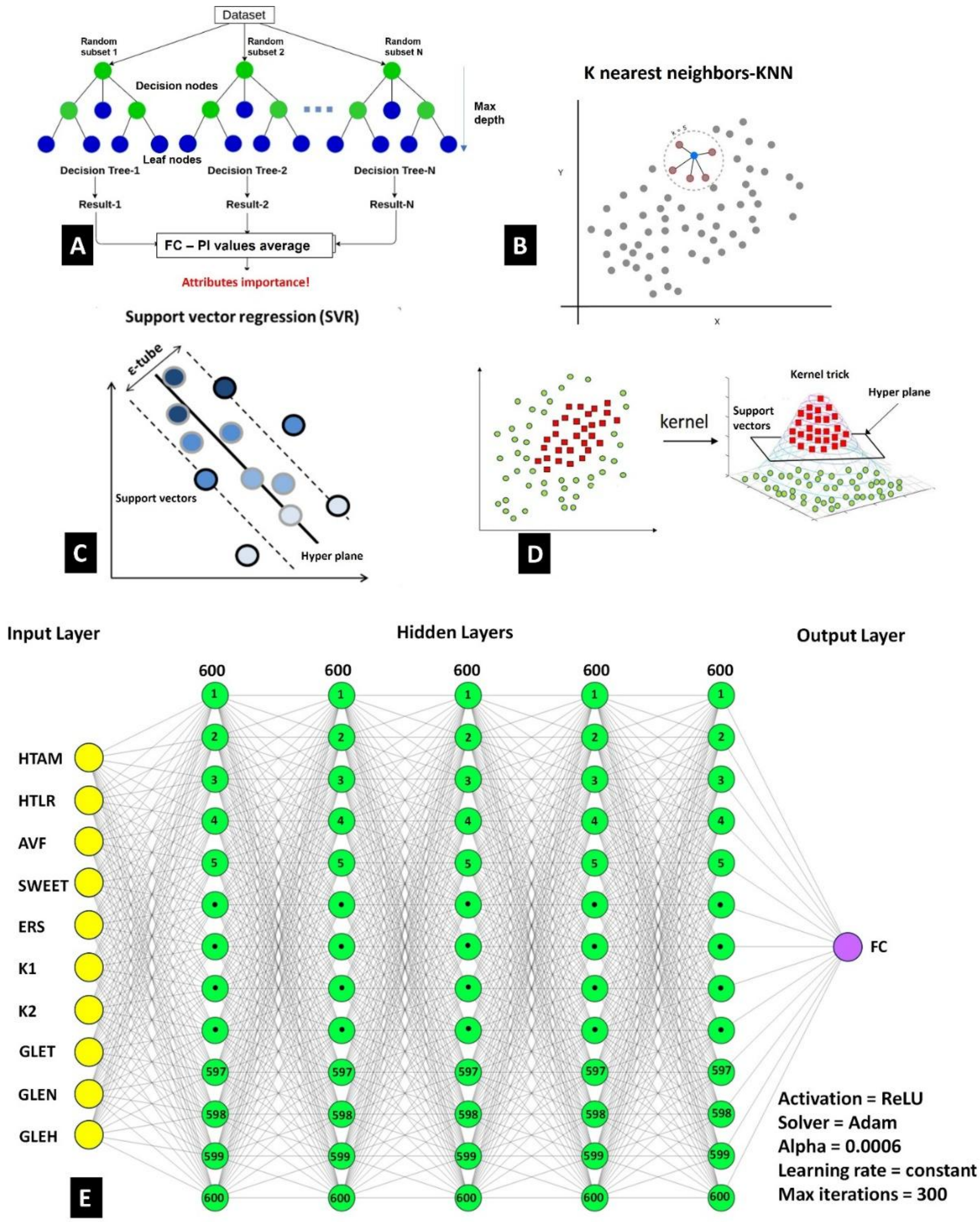

Figure 3.10. Some key aspects of the selected ML methods for this study. The illustrations here are simplified to 2D and 3D feature spaces. Keep in mind that in our case everything happens in an 11-dimensioned space. A) Illustration of random forest regression (modified from Goyal, 2022); B) Illustration of KNN regression: Low values of K can lead to overfitting, whereas very high values of K may not capture all the variance. C) Support Vector Machine regression: using support vectors (data points) and an epsilon tube, a hyper plane that best fits the data is built. In the case of non-linearly separable data, the kernel trick is used (D), where a higher dimensional space is emulated, so the hyper plane can be built (modified from Rodriguez Perez and Bajortah, 2022). E) Architecture and main hyperparameters of the MLP used in this study. At the input layer we have 10 seismic attributes at well tested intervals, 5 hidden layers with 600 neurons each (not all of them are shown in the figure), and the output layer with the prediction of flow capacity (FC).

Results

Table 3.2 summarizes the results of this study. It shows the main information about each one of the eleven supervised regression models generated with the four selected algorithms (RF, KNN, SVM, and MLP).

Supervised regression models

With ten input attributes, random forest supervised regression of FC led to a performance of 80% (16 out of 20) of correct predictions at the blind test wells positions. By correct prediction we mean: if the predicted value of FC, for example, lies in the same interval of the measured FC (for instance, predicted 5,000 mD.m versus measured 1,000 mD.m, both at the 0-10,000 interval – mild producers), it is considered a good prediction once wells with these FC values have similar PI values and production rates around 8,000 bpd (Figure 5). Moreover, the spatial distribution of the predicted values must follow a geological pattern (Figure 3.11A). With 80% performance of right predictions compared to the 5-fold GSCV R^2, which is 84% (Figure 3.11B), we can say that is a good model once it can generalize well the learning obtained during training. Figure 3.11A illustrates a 3D geobodies view of the predicted FC model using RF regression and ten input attributes. We used an opacity threshold of FC > 100,000 mD.m to render only the hyper flow units.

It is possible to separate the most productive reservoirs (hyper flow units at heavily silicified build ups and mounds) from good and mild producers from platform facies. The hyper productive build ups and mounds can produce, at the Mero Field, up to 48,000 bpd, while the higher and lower energy platform reservoirs produce 24,000 and 8,000 bpd respectively. At the inner platform region, these hyper flow units are interpreted as hydrothermal vents responsible for the strong silicification at build ups and mounds (Figure 3.5). This process is less pervasive at the platform facies because their lower permeability is related to lower depositional energy. So, the hydrothermal alteration is confined to faults and fractures. Wells drilled near such features, like Well C, often exhibit FC values higher than normal. Figure 11C shows the random forest attribute importance for FC prediction. The four most important attributes are AVF, K1, K2, and ERS. This is geologically accurate as the most productive reservoirs are related to heavily silicified and karstified build ups and mounds. These zones are often characterized by low frequency anomalies (due to karst attenuation of seismic signal), high K1-K2 values due to its convex shape, and low similarity values due to intense karstification, faulting, and fracturing. Figure 3.11D is a Shapley additive explanation chart (SHAP) for the same model. SHAP is a technique for explaining the predictions of machine learning models (Lundberg and Lee, 2017; Lundberg et al.,2018a, 2020; Molnar, 2021 and Lubo-Robles, 2022). SHAP is based on the concept of Shapley values, which quantify the contribution of each player (input seismic attributes) in a cooperative game (prediction of FC and PI). As random forest regression, SHAP also provides the individual importance of each input attribute to the prediction of FC and/or PI. As can be observed in Figures 3.11B and 3.11D, SHAP importance is very similar to random forest one, which indicates a strong geological consistency of the attribute's importance obtained with two very different methods.

For dimensionality reduction effect assessment on the regression models, we performed random forest regression of FC dropping attributes until the minimum number of six input features. As can be observed in Table 3.2, no significant drop in R^2 and blind test performance was observed. This is one of the most important findings of this study: with six attributes (AVF, K1, K2, ERS, HTAM, GLET) that can be generated from a simple full stack volume and random forest regression, we can predict FC at BVE reservoir with a performance of 80% in the blind test correct predictions (16 out of 20), if spectral balanced is applied, the performance increases to 85%. The deep learning model (MLP) is barely as good (75 % performance) as the random forest ones. Moreover, it is significantly more computationally expensive once MLP regression took 45 minutes for prediction in the blind test dataset, whereas RF took only 1.5 minutes (Table 3.2).

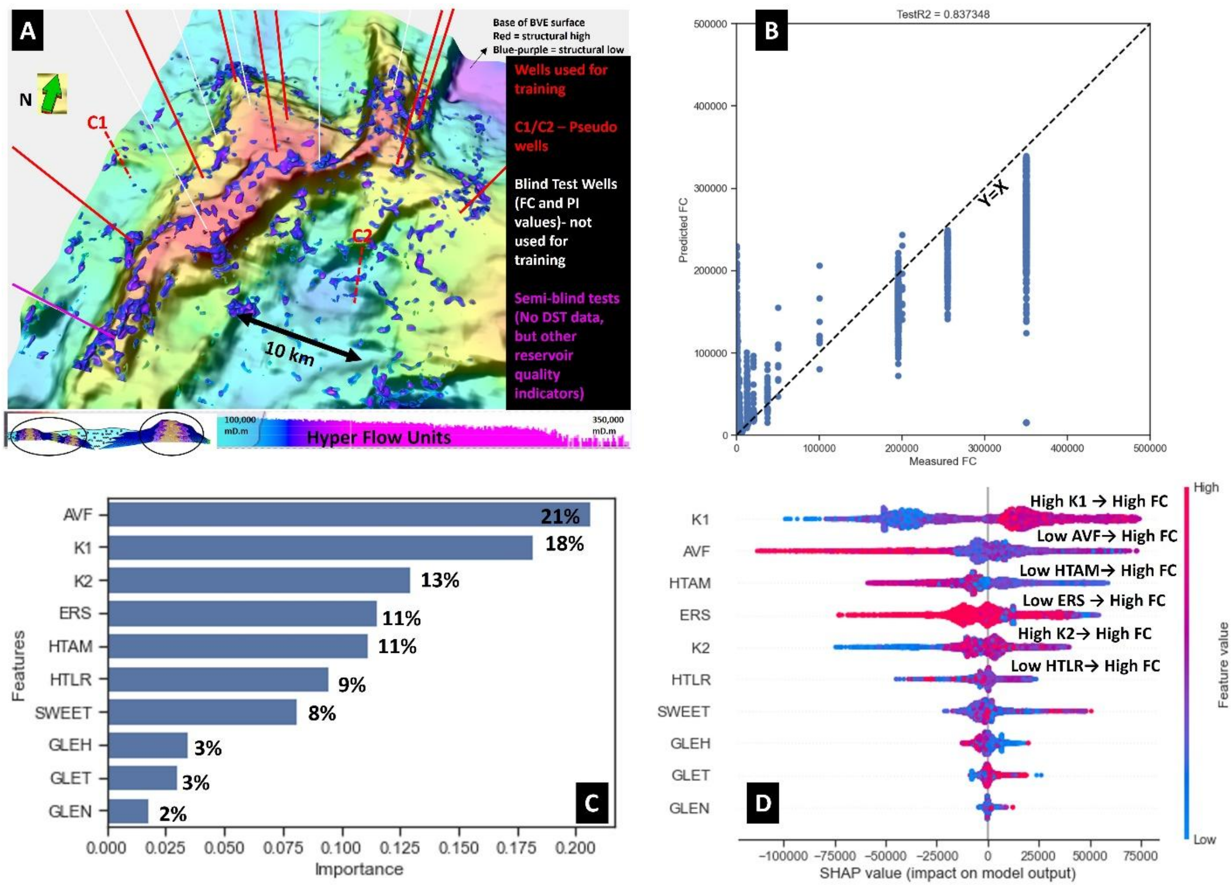

Figure 3.11. FC prediction using random forest (RF) and 10 attributes. A) Geobodies 3D view of the hyper flow units (FC > 100,000 mD.m). Despite not being restricted to depositional facies, these hyper flow units are equivalent mainly to silicified build ups and mounds. At both higher and lower energy platform facies, they are related to hydrothermal vents and tiny mounds. B) Plot of measured versus predicted FC from random forest training, test $R^2 = 0.84$. C) RF importance of the input attributes. D) SHAP explanation of this predictive model at training dataset.

Supervised ML Regression Model	Input seismic attributes	Predicted DST parameter	Algorithm	Best Hyperparameters after GSCV	Training dataset R^2 after GSCV	Blind test wells right predictions	Time for prediction (min)
RF_FC_10	AVF, K1, K2, ERS, HTAM, HTLR, SWEET, GLEH, GLET, GLEN	FC	RF	#Estimators = 500 Max depth = 10	0.84	16/20 (80%)	1.5
RF_FC_9	AVF, K1, K2, ERS, HTAM, HTLR, GLEH, GLET, GLEN	FC	RF	#Estimators = 500 Max depth = 10	0.83	16/20 (80%)	1.5
RF_FC_8	AVF, K1, K2, ERS, HTAM, GLEH, GLET, GLEN	FC	RF	#Estimators = 500 Max depth = 10	0.81	16/20 (80%)	1.5
RF_FC_6	AVF, K1, K2, ERS, HTAM, GLET	FC	RF	#Estimators = 400 Max depth = 10	0.82	16/20 (80%)	1.5
RF_FC_6_SB	AVF, K1, K2, ERS, HTAM, GLET	FC	RF	#Estimators = 400 Max depth = 10	0.86	17/20 (85%)	1.5
KNN_FC_6	AVF, K1, K2, ERS, HTAM, GLET	FC	KNN	K = 2	0.83	16/20 (80%)	1
SVM_FC_6	AVF, K1, K2, ERS, HTAM, GLET	FC	SVM	Kernel = RBF C = 1,000,000 Gamma = 100 e=0.1	0.82	11/20 (55%)	30
RF_PI_6	AVF, K1, K2, ERS, HTAM, GLET	PI	RF	#Estimators = 400 Max depth = 10	0.81	16/20 (80%)	1.5
KNN_PI_6	AVF, K1, K2, ERS, HTAM, GLET	PI	KNN	K=2	0.95	15/20 (75%)	1
SVM_PI_6	AVF, K1, K2, ERS, HTAM, GLET	PI	SVM	Kernel = RBF C = 1,000,000 Gamma = 100 e=0.1	0.98	9/20 (45%)	30
MLP_FC_10	AVF, K1, K2, ERS, HTAM, HTLR, SWEET, GLEH, GLET, GLEN	FC	MLP	# hidden layers=5 # neurons = 600 each Activation = ReLU Alpha = 0.0006 Max iterations = 300	0.91	15/20 (75%)	45

Table 3.2- Summary of the results of supervised machine learning (ML)regression of flow capacity (FC) and productivity index (PI) for BVE reservoirs at Mero Field and Central Libra block. Acronyms: GSCV (grid search cross validation), SB (Spectral Balancing); RBF (radial basis function), MLP (multi-layer perceptron), ReLU (rectified linear unit).

Vertical seismic sections

Figure 3.12 is a group of vertical seismic sections passing through a semi-blind test well (Well X1), which drilled an excellent reservoir quality carbonate build up with strong hydrothermal silicification. Figures 3.12A to 3.12F contain the selected input attributes for random forest regression of FC and PI (Figures 3.12 G and 3.12 H). In Figure 3.12I, we have Well X1 composite log at the BVE reservoir with lithology, nuclear magnetic resonance porosity (NMR), and lithogeochemical spectroscopy log, which is useful for hydrothermal zones identification by measuring the relative concentration of major chemical elements (Ca, Fe, Mg, Si, H, and Cl) (Griffiths, 2010). At this well position, the entire BVE reservoir is heavily silicified and very porous and permeable as indicated by mini-DST data. It was a pre-existing karst system at a stromatolitic build up that was amplified by silica-rich hydrothermal fluid that percolated preferentially through the cave system at the build up and mound positions. At the very top of the reservoir, there is a 10 m thick alkaline sill intrusion (Santonian age). Some of the above-mentioned features can be seen in some seismic attributes. The karst system features, which are often associated with deep faults and fractures at build up positions, can be seen at AVF, ERS, and GLET attributes (Figures 3.12A, 3.12D, and 3.12F). These attributes are suitable to portray karstified zones (centimetric to decametric cave systems). In the seismic scale (> 50m in the case) the karst zones appear as low values of AVF (seismic attenuation due to caves), ERS, and GLET (caves and fractures). Hydrothermal vents at lower energy facies can also be seen in these attributes, whereas the igneous intrusion at the top of the reservoir can be recognized at HTAM and GLET attributes. These features that are illustrated in the reservoir model of Figure 3.12J are well represented in the supervised regression of FC and PI models where the hyper flow units are located over the silicified build up drilled by Well X1 and hydrothermal vents at its flanks. The igneous body was correctly predicted as a very low both FC and PI flow units (Figures 3.12G and 3.12H).

Figure 3.13 presents other vertical sections at some key well positions. In Figure 3.13A, we have the same FC random forest regression model with six attributes of Figure 3.12 as a transparent overlay over the HTAM attribute. Wells I and J were used for training and Well BT17 is a blind test, all of them drilled very productive build up features. The FC predictions (hyper flow units) at both wells I and J are consistent with image logs at the silicified and karstified reservoirs characterized by high NMR porosity and silica content (Figure 3.13B). Well I has a metric scale cave system and the FC values are 255,000 mD.m. At well J, the cave system is centimetric and the FC drops to less than 200,000 mD.m, demonstrating the effect of karstification intensity on reservoir productivity. The blind test BT17 in Figure 3.13A is the best producer at the BVE reservoir at Mero Field (48,000 bpd) and the FC of 195,000 mD.m was correctly predicted by our random forest regression using six attributes. It is important to observe that both wells BT17 and J were drilled at the same build up feature which is displaced by a normal fault. In Figure 3.13C, we can see the same RF regression of FC but with the six input attributes derived from a spectrally balanced full stack volume (higher resolution). At platform facies positions like Well G, flow units tend to be more layered driven by the nature of the reservoir, which is the intercalation of spherulitites, grainstones, and microbial mats. By comparing the composite logs of Well BT17 (Figure 3.13B) and Well G (Figure 3.13D), we can see the difference in terms of porosity distribution and silica content. The "spiky" aspect of the NMR log is related to intense fracturing and karstification indicated by the image log, where a decametric cave system can be observed

(Figure 3.13B). This kind of reservoir has the highest FC and PI values in Mero Field and can produce up to 48,000 bpd. As Well G was drilled in a lower energy platform area, NMR porosity is lower than at build ups, and the low silica content is related to the terrigenous content of BVE fine-grained carbonates (Figure 3.13D). This lower energy facies reservoir has typical FC and PI values lower than 10,000 mD.m and 100 $m^3/d.kgf.cm^2$, respectively. Above normal FC values measured at well G (24,000 mD.m) can be explained by the presence of near-fault hyper flow units related to hydrothermal vents and/or minor mounds (Figures 3.13A and 3.13C).

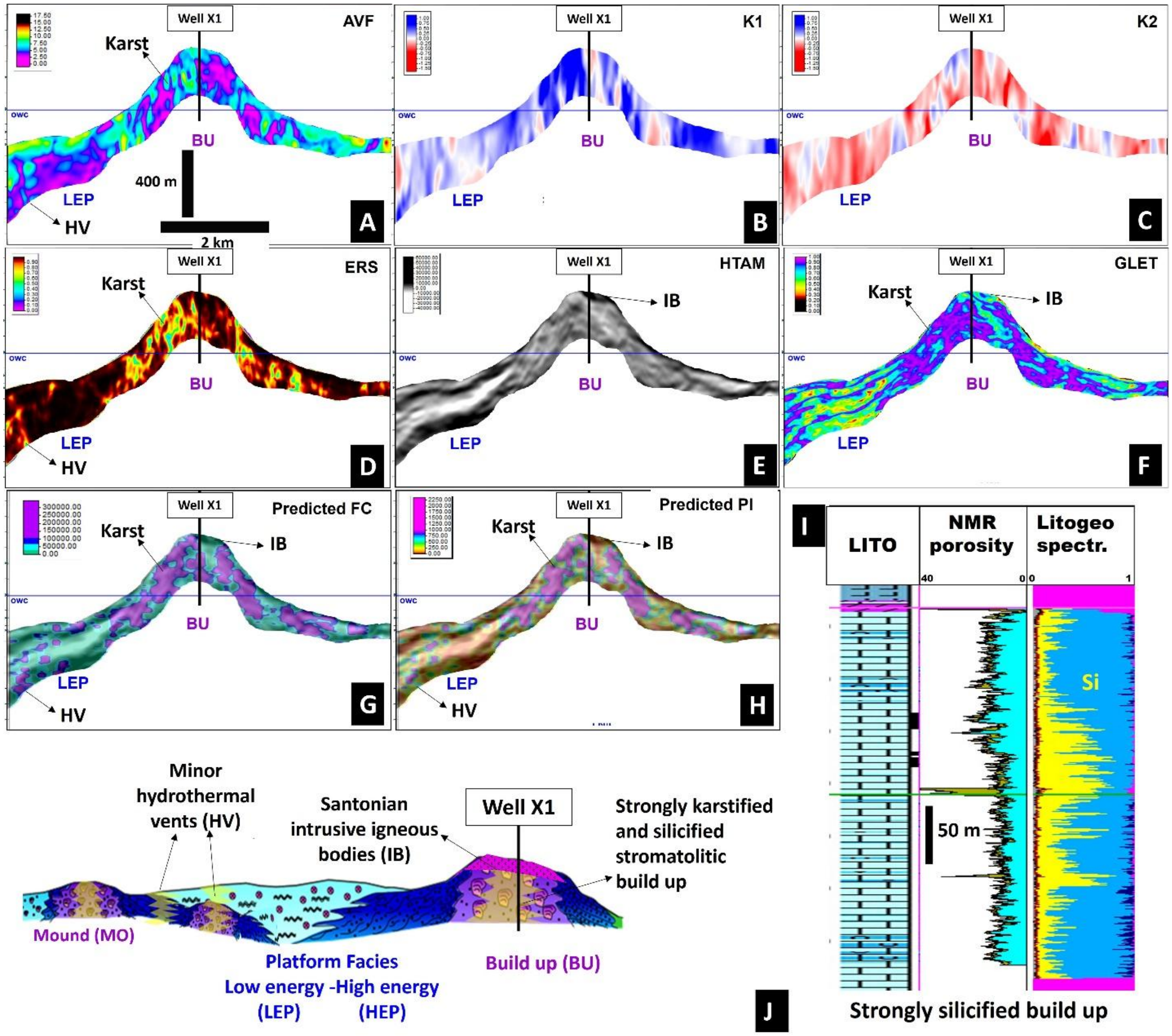

Figure 3.12. Six of the input attributes at a semi-blind test well location (X1) and RF regression models of FC and PI at BVE Formation. This well penetrated a very high porosity and permeability reservoir related to highly silicified stromatolitic build up. This feature and others can be observed at both input seismic attributes and predicted models of FC and PI. A) Average frequency; B) Maximum positive curvature; C) Maximum negative curvature; D) Energy ratio similarity; E) Hilbert transform of amplitude; F) GLCM entropy; G) Predicted flow capacity; H) Predicted productivity index. I) Composite log of the BVE interval with lithology, NMR porosity, and lithogeochemical spectroscopy log showing silicon concentration (Si). J) Illustration of Well X1 position according to the BVE depositional model (modified from Mizuno et al., 2018). Geologic features acronyms: BU (Build ups), MO (Mounds), HEP (high energy platform facies), LEP (low energy platform facies), IB (igneous bodies), HV (hydrothermal vents).

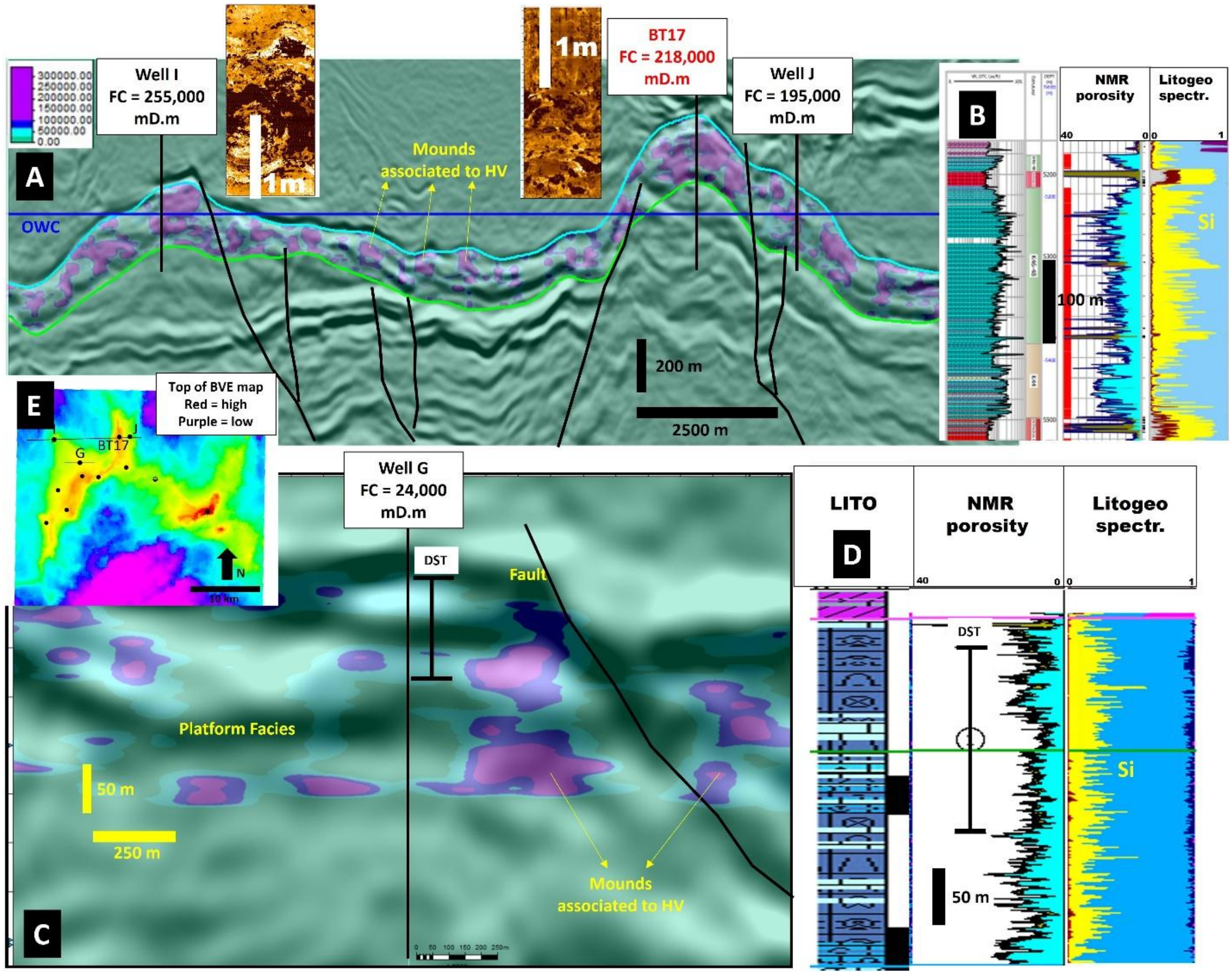

Figure 3.13. RF regression models of FC using six attributes (AVF, K1, K2, ERS, HTAM, GLET) as a transparent overlay on input seismic amplitude (full stack RTM in depth domain). A) EW vertical section passing through wells I (training), BT17 (blind test), and J (training). Close to wells I and BT17, there are image logs showing the cave system which is responsible for the observed very high productivity; OWC= oil water contact; B) Composite log of Well BT17 with lithology, NMR porosity and lithogeochemical spectroscopy; C) EW vertical section passing through Well G (training); D) Composite log of Well G. This model was predicted using the same attributes as in Figure 13A, but after spectral balancing of original seismic data.

Discussion

After blind test validation, when comparing random forest regression with other shallow learning algorithms, K-nearest neighbors showed similar performances, but support vector machine models were more prone to overfitting, once a performance of 55% was observed at blind test wells, while the GSCV R^2 of SVM regression of FC is 0.82. This can be related to the selected kernel type (radial basis function), which is prone to fit very complex models to the training dataset and, thus, overfitting it. Support vector machine prediction took 30 minutes, while random forest and K-nearest neighbors took only 1.5 and 1 minute respectively (Table 3.2). The attribute combination used for supervised regressions of FC and PI performed well, as they are geologically meaningful and mathematically independent. From the tested supervised ML algorithms, we found that random forest regression is the best one once it led to the most accurate predictions of FC and PI at the blind test wells positions (80 – 85%). The most important finding is that without any seismic inversion attribute and even AVO, it is possible to seismically predict FC and PI at the BVE reservoir in Mero Field and CLAP area with blind test performance of up to 85%. When compared to the deep learning models (MLP), random forest is better once it has the highest blind test performances at a very low computational cost (1.5 minutes for prediction against 45 minutes for MLP– Table 3.2). Figure 3.14 illustrates a more regional 3D view of random forest regression predicted flow capacity with six attributes (AVF, K1, K2, ERS, HTAM, and GLET). At CLAP, an underexplored area, there is a semi-blind test well (X5) which was not used for any training. At this well position, the BVE reservoir is excellent (silicified build up) and was correctly predicted. Therefore, this model is relatively well calibrated considering its blind test performance at Mero Field (80%) and correct predictions at the CLAP area. Around Well X5, several geobodies of hyper flow units can be observed above and below the oil-water contact (OWC) and could be considered future targets for drilling. As mentioned before, the hyper flow anomalies located in the structural lows are related to minor mounds along hydrothermal vents close to deep faults within distal facies (likely chemosynthetic communities). At the higher and lower energy platforms, these anomalies are related to hydrothermal alteration close to the vents and related minor mounds. At Mero Field, in the lower energy platform, wells located around these anomalies often exhibit above-normal productivity. Nevertheless, once the blind test performance is high (80% or 16 out of 20 wells), one fifth (4 wells) of the predictions are wrong (mostly overpredictions), this is a limitation of the selected machine learning methods. So, some of the undrilled hyper flow anomalies can be false. Despite the remaining low risk (20%), hyper flow anomalies underneath OWC could be considered as future targets for CO_2 injection, for example.

It is worth mentioning that wells like the most productive ones at the Mero Field (silicified and karstified build ups) can produce up to 48,000 bpd (BVE reservoir only). With such wells, only four pairs of injectors/producers are enough to fulfill a 180,000-bpd floating, production, storage, and offloading unit (FPSO). So, if it is possible to predict and drill these high-yield wells as soon as possible in the E&P project timeline, a lot of economic value will be added to the project by accelerating and, thus, optimizing commercial production.

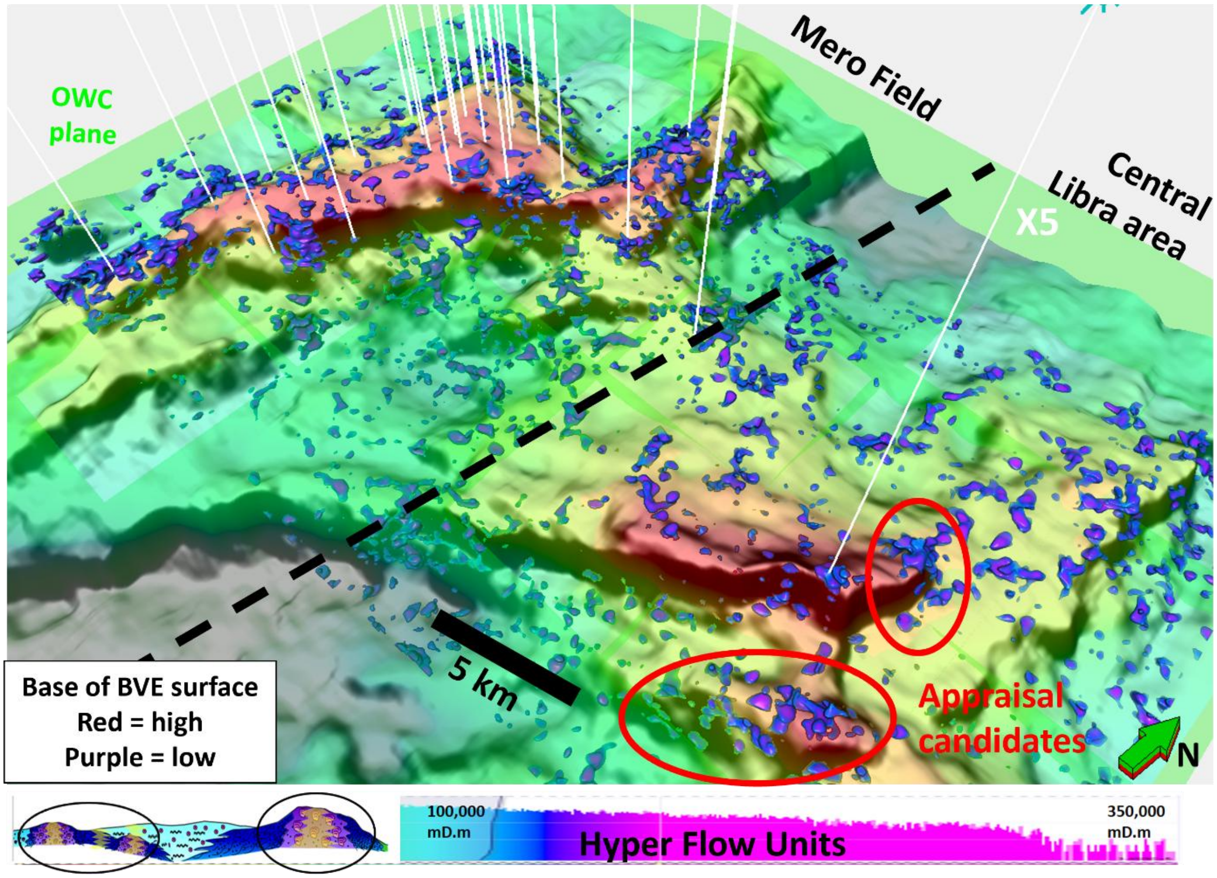

Figure 3.14. 3D view of random forest regression of FC at Mero Field and CLAP area (geobodies) over the base of BVE surface. The hyper flow units related to sweet spot reservoirs (silicified build ups, mounds, and hydrothermal vents) are rendered with an opacity threshold of 100,000 mD.m. At the Central Libra area (CLAP), features like that drilled by Well X5 (sweet spot reservoir) that are above the oil-water contact identified only at the Mero Field (OWC- green plane) could be considered as future targets for oil exploration. Features below OWC can be considered as future injection targets. Hyper flow units away from structural highs are related to either minor mounds or higher permeability, altered platform facies close to hydrothermal vents. There is still a misprediction risk once at 4 out of 20 blind test wells (20%), the FC values were overpredicted (false anomalies).

Conclusions

We created a reservoir de-risking tool that, despite being not perfect (20 % of wrong predictions during blind test validation), can be used for carbonate reservoir productivity prediction at under-appraisal, under-development fields, and underexplored areas. With the supervised machine learning methods presented here, mainly random forest, we could unravel hidden relationships between simple seismic attributes and DST/PLT flow units. We built both FC and PI models that fit well the training data and predicted these dynamic parameters with high performance (80 – 85%) on the blind test dataset. Both random forest and SHAP have explained the prediction models by ranking the attribute's importance. Furthermore, SHAP has described how each one contributes to FC prediction: the hyper flow units related to very productive silicified build ups and mounds are related to low AVF (seismic attenuation due to karstification), high K1-K2 due to their convex shape, low ERS (fault-fractures-karst) and both low HTAM and HTLR related mainly to karst giga-porosity and silicification. The highest performance at blind test well positions (85%) was obtained with random forest regression using six attributes that can be derived from a spectrally balanced full stack volume (AVF, K1, K2, ERS, HTAM, and GLET). Neither seismic inversion nor AVO data were needed to reach such a high blind test performance. This can help to speed up the decision-making process when these data are not available. At CLAP, an underexplored area, our models correctly predicted the excellent reservoirs of semi-blind test Well-X5 and identified several targets for potential appraisal well locations. Therefore, this attribute combination is suitable for BVE reservoir productivity prediction in areas at earlier exploration stages, where seismic inversion data are not reliable and AVO data are not always available. For the purposes described in this paper, deep learning predictive models barely provided similar results as shallow learning, but at a much higher computational cost. The next research steps will focus on machine learning transfer from Mero Field to other seismic surveys in areas with geological similarities and challenging problems such as low productivity reservoirs and carbonate versus volcanic mounds discrimination.

Acknowledgements

We would like to thank the Libra Consortium (Petrobras, Shell Brasil, TotalEnergies, CNPC, CNOOC, and PPSA), and the Brazilian National Petroleum Agency (ANP) for releasing the dataset used in this work. Thanks to SLB for donations of the Petrel software to the University of Oklahoma (OU). Thanks to the open-source coding providers (Python, NumPy, Pandas, and Scikit-Learn). Special thanks to Petrobras for sponsoring the first author's PhD research done so far with the AASPI research group at OU. Many thanks also to other OU colleagues for suggestions and ideas exchange, especially Dr. David Lubo-Robles. Special thanks to Dr. Deepak Devegowda (OU Petroleum Engineer Professor) for the very insightful lectures and comments on data analytics and machine learning. We are also very grateful to Conrado Keidel (Petrobras Petroleum Engineer), for insightful discussions and for reviewing parts of the manuscript.

References

Abdulaziz, M. A., Mahdi, H. A, Sayyouh, M.H. 2019. Prediction of reservoir quality using well logs and seismic attributes analysis with an artificial neural network: A case study from Farrud Reservoir, Al-Ghani Field, Libya, Journal of Applied Geophysics, Volume 161, Pages 239-254, https://doi.org/10.1016/j.jappgeo.2018.09.013.

Alfarraj, M., Alregib, G. 2018. Petrophysical-property estimation from seismic data using recurrent neural networks, SEG Technical Program Expanded Abstracts: 2141-2146. https://doi.org/10.1190.

Anifowose, F., Abdulraheem A., Al-Shuhail, A. 2019. A parametric study of machine learning techniques in petroleum reservoir permeability prediction by integrating seismic attributes and wireline data, Journal of Petroleum Science and Engineering, Volume 176, Pages 762-774, https://doi.org/10.1016/j.petrol.2019.01.110.

Ballinas, M. R., Bedle, H., Devegowda, D. 2023. Supervised machine learning for discriminating fluid saturation and presence in subsurface reservoirs, Journal of Applied Geophysics, Volume 217, 105192, ISSN 0926-9851, https://doi.org/10.1016/j.jappgeo.2023.105192.

Bourdet D., 2002. Well-test Analysis: The use of advanced interpretation models, Elsevier.

Brunton, S.L., Kutz, J.N.2019. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press, Cambridge.

Carlotto M.A, Silva R.C, 2018. Libra: A Newborn Giant in the Brazilian Pre-Salt Province. AAPG Annual Convention and Exhibition, Expanded Abstracts, Salt Lake City.

Chopra, S., Marfurt, K.J., 2007. Seismic Attributes for Prospect Identification and Reservoir Characterization: Society of exploration geophysicists.

Cortes, C. and Vapnik, V. 1995. Support-Vector Networks. Machine Learning, 20, 273-297. http://dx.doi.org/10.1007/BF00994018.

Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P.J., Levitt, P.R., 1994. Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique: Geophysics, 59, 1362–1376.

Fix, E. and Hodges, J.L. (1951) Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.

Fukami, K., Fukagata, K., Taira, K. 2020. Assessment of supervised machine learning methods for fluid flows. Theoretical and Computational Fluid Dynamics. 34. 10.1007/s00162-020-00518-y.

Goyal, C. 2022. 25 Questions to Test Your Skills on Random Forest Algorithm. https://www.analyticsvidhya.com/blog/2021/05/bagging-25-questions-to-test-your-skills-on-random forest-algorithm.

Griffiths R., 2010. Ecoscope User's Guide. Schlumberger.

Ho, T.K. 1995. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada, pp. 278-282 vol.1, doi:10.1109/ICDAR.1995.598994.

Hosseini, M., Riahi, M. A., & Mohebian, R. 2019. A Meta attribute for reservoir permeability classification using well logs and 3D seismic data with probabilistic neural network. Bollettino di Geofisica Teorica ed Applicata. 60. http://dx.doi.org/10.4430/bgta0257.

Kangxu, R., Oliveira, M.J., Zhao, J., Zhao, J., Oliveira, L., Rancan, C., Carmo, I., Deng, Q. 2019. Using Wireline Logging and Thin Sections to Identify Igneous Contact Metamorphism and Hydrothermal Influence on Presalt Limestone Reservoirs in Libra Block, Santos Basin. 10.4043/29818-MS.

Lima, B. E., Tedeschi, L., Pestilho, A. L., Santos, R., Vazquez, J., Guzzo, J., De Ros, L. 2020. Deep-burial hydrothermal alteration of the Pre-Salt carbonate reservoirs from northern Campos Basin, offshore Brazil: Evidence from petrography, fluid inclusions, Sr, C and O isotopes. Marine and Petroleum Geology. 104143. 10.1016/j.marpetgeo.2019.104143.

Lubo-Robles, D., D. Devegowda, V. Jayaram, H. Bedle, K. J. Marfurt, and M. J. Pranter. 2022. Quantifying the sensitivity of seismic facies classification to seismic attribute selection: An explainable machine-learning study: Interpretation, 10, no. 3.

Lundberg, S., S. Lee. 2017. A unified approach to interpreting model predictions: Proceedings of the 31st Conference on Neural Information Processing Systems, 4765–4774.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. Lee. 2020. From local explanations to global understanding with explainable AI for trees: Nature Machine Intelligence, 2, 56–57, doi: 10.1038/s42256-019-0138-9.

Lundberg, S. M., G. G. Erion,S. Lee. 2018a. Consistent individualized features attribution for tree ensembles: arXiv preprint arXiv:1706.06060.

Maas, M. V. R., Bedle, H., Matos, M. C. 2023.Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water Aptian pre-salt data. Marine and Petroleum Geology, Volume 151, https://doi.org/10.1016/j.marpetgeo.2023.106199.

Mitra, P. P., Kharak S. 2001. Synergy through Seismic Attributes in Reservoir Management- a Case Study. Paper presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia, doi: https://doi.org/10.2118/68648-MS.

Molnar, C. 2021. Interpretable machine learning: A guide for making black box models explainable, https:// christophm.github.io/interpretable-ml-book/shap.html, accessed 11 August 2021.

Moreira, J.L.P., Madeira, C.V., Gil, J.A., Machado, M.A.P. 2007. Bacia de Santos. Rio de Janeiro. Boletim de Geociências da Petrobras, v. 15, n. 2, 531-549.

Oliveira L. C., Rancan, C. C. (2018): Ocorrência de condutos hidrotermais (hydrothermal vents) na seção aptiana da Bacia de Santos. In Congresso Brasileiro de Geologia, 49, Rio de Janeiro, 2018, p. 2008.

Oliveira, L.C., Rancan, C.C., 2019. Sill emplacement mechanisms and their relationship with the Pre- Salt stratigraphic framework of the Libra Area (Santos Basin, Brazil). LASI VI-The physical geology of subvolcanic systems: laccoliths, sills and dykes. Malargüe, Argentina.

Oliveira, Leonardo Costa de and Penna, Rodrigo Macedo and Rancan, Cristiano Camelo and Carmo, Isabela de O. and Marins, Gabriel Medeiros, Seismic Interpretation of the Mero Field Igneous Rocks and its Implications for Pre- and Post-Salt Co2 Generation – Santos Basin, Offshore Brazil. Available at SSRN: https://ssrn.com/abstract=4394061 or http://dx.doi.org/10.2139/ssrn.4394061

Penna, R., Lupinacci, W. 2020. Decameter-Scale Flow-Unit Classification in Brazilian Presalt Carbonates. SPE Reservoir Evaluation & Engineering. 23. 1-20. 10.2118/201235-PA.

Poros, Z., Jagniecki, E., Luczaj, J.,Kenter, J., Gál, B., Correa, T., Diniz Ferreira, E. L., Heumann, M., Elifritz, A., Johnston, M., Mcfadden, K., Matt, V. 2017. Origin of Silica in Pre-Salt Carbonates, Kwanza Basin, Angola. 10.13140/RG.2.2.31863.42405.

Raeesi, M., Moradzadeh, A., Ardejani, F.D., Rahimi, M. 2012. Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks, Journal of Petroleum Science and Engineering, Volumes 82–83, Pages 151-165, https://doi.org/10.1016/j.petrol.2012.01.012.

Rodríguez-Pérez, R., Bajorath, J. 2022. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *J Comput Aided Mol Des* **36**, 355–362. https://doi.org/10.1007/s10822-022-00442-9.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6):386.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323:533–536.

Sartorato, A.C.L, Mizuno, T.A., Paegle, J.S., Pozzi, R.P.S, Zanatta, A.S., Rosseto, J.A., Oliveira, L.C., Penna, R.M., 2018. Modelagem geológica do reservatório do Campo de Mero. XI Seminário de Interpretação Exploratória. PETROBRAS. Rio de Janeiro.

Sartorato, A.C.L., Tonietto, S.N., Pereira, E., 2020. Silicification and Dissolution Features in the Brazilian Presalt Barra Velha Formation: Impacts in the Reservoir

Quality and Insights for 3D Geological Modeling. Rio Oil & Gas, Rio de Janeiro, Brazil. https://doi.org/10.48072/2525-7579.rog.2020.068.

Stica, J. M.; Zalan, P. V.; Ferrari, A. L., 2014. The evolution of rifting on the volcanic margin of the Pelotas Basin and the contextualization of the Paraná-Etendeka LIP in the separation of Gondwana in the South Atlantic. *Marine and Petroleum Geology*, [Amsterdam], v. 50, p. 21.

Taboada, R., Condat, P., Corsini, V., Mir, E., Conti, J., Fortunato, G., Villivar, O., List, D., Stancel, S., and L. Jones. 2001. El Tordillo Reservoir Static Characterization Study: El Tordillo Field, Argentina. Paper presented at the SPE Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, March 2001. doi: https://doi.org/10.2118/69660-MS.

Tritlla, J. & Loma, R., Sanchez Perez-Cejuela, V., Cerdà, M., Wloszczowski, D., Perona, R., Algíbez, J., Caja, M. A., Levresse, G., Bonillo, P., Tiwary, D. ,2013. Volcanism, alkalinity and hydrothermalism in offshore Brazil: unraveling the Pão de Açúcar reservoir. Repsol E&P Conference Madrid, Spain.

Zalan, P. V. ,2014. Comparisons between the Deep Crustal Structure of Magma-Poor and Volcanic Passive Margins. In: Atlantic Conjugate Margins Conference, 4. St. John's, Canada. *Extended Abstracts*… St. John's, Canada: [S.n.], 2014. p.75–78.

Chapter 4- Transfer learning from seismic data for predicting reservoir productivity and lithologies for upstream optimization

*This study was submitted (still under review) for publication SEG Interpretation journal as

Maas, M. V. R., Bedle, Maas, E.C.A.S. 2025. Transfer learning from seismic data for predicting reservoir productivity and lithologies for upstream optimization

Abstract

While seismic inversion is an excellent tool for reservoir characterization, its reliability hinges on extensive well control that is both area-specific and non-transferable between surveys. This limitation becomes particularly acute in deepwater environments, where developing a dependable inversion cube can span more than a decade. To address these constraints, we present a transfer learning approach for reservoir prediction that leverages well dynamic data and post-stack seismic attributes available during initial exploration phases. We used data from a known oilfield (pre-salt Mero oilfield, Brazil offshore) covered by narrow azimuth seismic data to train and validate machine learning models, mainly self-organizing maps and random forest regression. From a depth-migrated amplitude volume, we derived nine post-stack attributes (complex trace, geometric, and voxel-based texture).

After data similarity checks across different seismic surveys using principal component analysis, exploratory data analysis, and feature engineering, we applied transfer learning to other areas with a few wells that were used as blind tests for models' validation. Next, we ran self-organizing maps and random forest regression of flow capacity (trained with Mero data) in two other surveys hundreds of kilometers distant from the Mero field: Bacalhau and Lapa oilfields, there is no similar approach in the literature. At the Bacalhau field, we differentiated a non-productive volcanic mound from hyper-productive carbonate mounds (blind test performance = 85%). At the Lapa field, we correctly predicted the lower reservoir flow capacity at the perforated zones (blind test performance = 75%).

We created a disruptive method that can transfer machine learning from well-characterized areas to others with poor well control and make accurate reservoir productivity predictions. Therefore, it can be used as a powerful reservoir de-risking tool in upstream phase projects. Likewise, it can be applied to optimize any other subsurface projects like CCUS, geothermal, and hydrogen storage in the energy transition context.

Introduction

One of the main challenges when working with seismic attributes-based machine learning (ML) subsurface geology prediction, is to transfer the learning obtained at the training dataset seismic survey to other surveys located dozens to even hundreds of kilometers away, acquired at different times, with distinct acquisition/processing technologies and parameters that led to totally different seismic amplitude ranges. Bäck

(2019), Wong et al., (2022), and Ehrig et al., (2024) stated that one critical aspect of transfer learning (TL) is the similarity between the source (training) and target (testing) datasets. In geology and geophysics (G&G) tasks, this similarity is primarily related to the geological semblance between the reservoir across source and target datasets either in the same sedimentary basin, one of its sectors, or even at different basins. This similarity must be in terms of geological age, depositional environment, and lithological facies distribution within the reservoir. The greater the similarity, the more feasible and successful will be the prediction performance across different seismic surveys. Several authors used TL for different G&G tasks such as seismic fault detection (Cunha et al., 2020), seismic images segmentation (Civitarese et al., 2018), earthquake seismology (Chai et al., 2020), and reservoir property prediction (El Zini et al., 2015; Dong et al., 2021; Yang et al., 2023; Huang and Tian, 2024, and Yang et al., 2024). However, those analyses used either well logs, synthetic seismic data, or the same seismic survey data, without dealing with the complexity of additional and different seismic surveys. In this pioneering and unique work, we explore the application of TL using pre-trained and validated supervised and unsupervised machine learning models of carbonate reservoir productivity using post-stack seismic attributes and production data from a well-known producing oil field as training data set (Mero oilfield, source area). We also explain how to make accurate predictions in other assets covered by different seismic surveys hundreds of kilometers apart (target areas), but still in the same basin with high geological similarities. However, there are also important lithological discrepancies between the fields related to hydrothermally modified carbonates and the presence of non-productive volcanic mounds that are like hyper-productive carbonate mounds in terms of seismic amplitudes. This is a continuation of the works of Maas et al. (2023) and Maas et al. (2024) that used post-stack seismic multi-attributes as input for self-organizing maps (SOM) clustering and random forest regression (RFR) among other ML techniques for carbonate reservoir productivity prediction using drill stem test (DST) data for training and validation of the RFR flow capacity (FC or kh, permeability*thickness) models at the Mero field and adjacent areas covered by the same seismic survey. Herein, we used parameters from these pre-trained models to make predictions in other datasets (different seismic surveys) that have DST data that were used as blind tests for the predictive model's validation. We also demonstrate how feature engineering was applied to overcome local differences in the reservoir's geology for TL optimization.

This paper is structured as follows: First, we describe the geological background and similarities between the Mero, Bacalhau, and Lapa fields, focusing on the key characteristics of the Barra Velha Formation reservoir. Next, we detail our dataset and methodology, including the selection and computation of seismic attributes, and our approach to feature engineering for transfer learning optimization. We then present the results of applying transfer learning from the Mero field to both Bacalhau and Lapa fields, demonstrating successful prediction of flow capacity and facies distribution. Finally, we discuss the implications of our results and potential applications beyond hydrocarbon exploration, including applications for CCUS, geothermal energy, and hydrogen storage projects.

Area location and Geological Background

We trained and tested our ML algorithms using seismic and well data from the Brazilian pre-salt province Mero Field (training dataset, source area) and Lapa and Bacalhau Fields (blind test validation datasets, target areas) (Figure 4.1). In terms of geology, all three fields are in the prolific Santos Basin and the main reservoirs are the Late Aptian lacustrine carbonates of Barra Velha Formation (BVE). More detailed studies on the stratigraphy, petroleum systems, and tectonostratigraphic evolution of the Santos Basin were carried out by Moreira et al. (2007), Zalan (2014), Stica et al. (2014), and Zalan (2024). The BVE reservoirs are dominantly microbially induced fascicular calcite aggregate crusts that occur as mounds and build ups (reef-like features) close to deep faults inherited from basement horsts (Figure 4.2). During subaerial exposure times, the mounds and build ups were eroded and re-deposited as rudstones and grainstones (composed of mound fragments) as gravitational and wave reworking deposits, respectively. At the lower energy platforms, there are layered sequences composed of in situ microbial mats (microbiolitic ramps), spherulites and Mg-rich mudstones, which dominate in the distal parts (Gomes et al., 2019). After deposition, those carbonates experienced multiple episodes of diagenesis such as burial dolomitization, strong hydrothermal alteration along reactivated faults followed by carbonate dissolution by corrosive silica rich fluids, ending up in either micro or mega quartz precipitation (Figure 2) (Tritlla et al. 2013; and Poros et al. 2017; Oliveira and Rancan, 2018; Kangxu et al., 2019; Oliveira and Rancan, 2019; Lima et al.,2020; Correa, 2023, and Oliveira et al., 2024). In Figure 4.2, we can see that geology across the three different fields is roughly the same with some local differences that will be further addressed in the following sections. The most important aspect is that in the BVE formation, because of the iteration of depositional, diagenetic, tectonic, and hydrothermal processes, there is a dual perm-porosity system. One end member is characterized by lower permeability matrix-dominated layered reservoirs with very low primary porosity and permeability related to fine-grained either low energy platform or distal carbonates. In the other extreme, in the carbonate mounds and build ups, there is a superimposition of epigenic and hypogenic karstification. The original cave system was amplified by corrosive silica-rich hydrothermal fluids with quartz precipitation at the end of the process (Correa, 2023). As a result, there is a net permeability gain (Mimoun and Fernandez-Ibanez, 2023). In between these end members, there is a depositional permeability obliterated by dolomitization that was locally dissolved along reactivated faults, where corrosive, silica-rich fluids have circulated followed by quartz precipitation. The overall result is a net permeability loss (Figure 4.3). According to the studies of Correa (2023) on deposition and hydrothermal alteration absolute dating (U-Pb) of BVE reservoirs, the chronological sequence of events is: carbonate deposition (Aptian-116 Ma), early dolomitization (Aptian-116 Ma), burial dolomitization (Albian-101 Ma), hydrothermal dolomitization along reactivated faults (Cenomanian-95 Ma), dissolution along faults by silica rich fluids, and, mega quartz-calcedony precipitation (no absolute age constraint, only petrographic evidence). All these events predated the oil charge, which took place at the Cretaceous-Paleogene boundary (65 Ma, Yan et al., 2024).

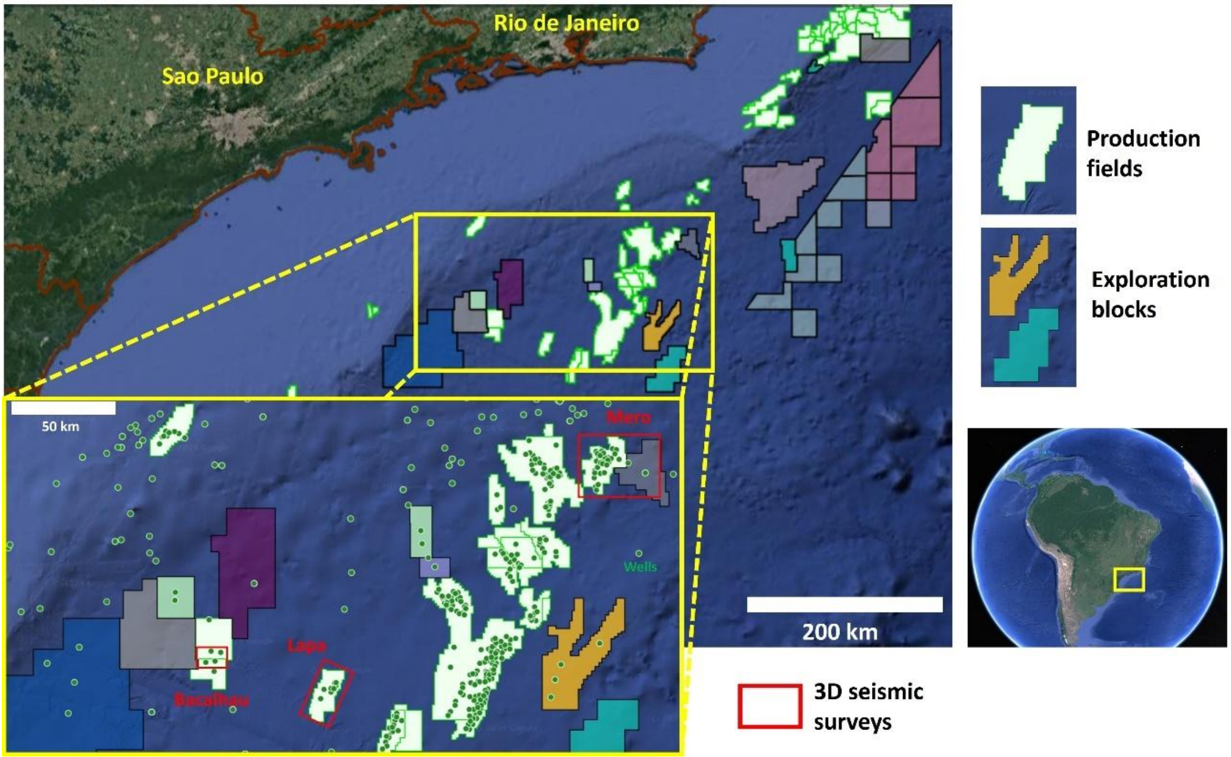

Figure 4.1- Location map of our studied areas in Brazil's offshore Santos. As a training dataset, we used well and seismic data from the Mero field. As blind validation of transfer learning, we used seismic and well data from Bacalhau and Lapa fields (map source: https://geomaps.anp.gov.br)

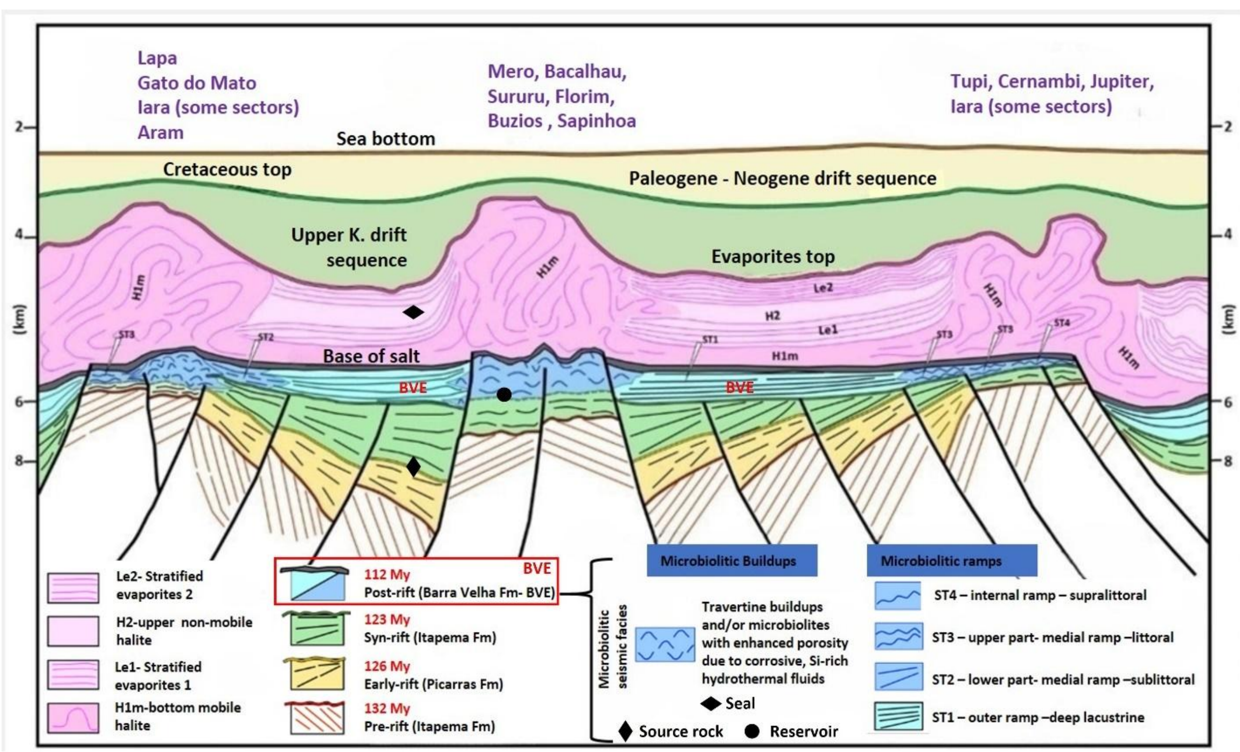

Figure 4.2 – Schematic geological section showing the general similarities at the Barra Velha (BVE) reservoir between Mero, Lapa, and Bacalhau fields (modified from Zalan, 2024).

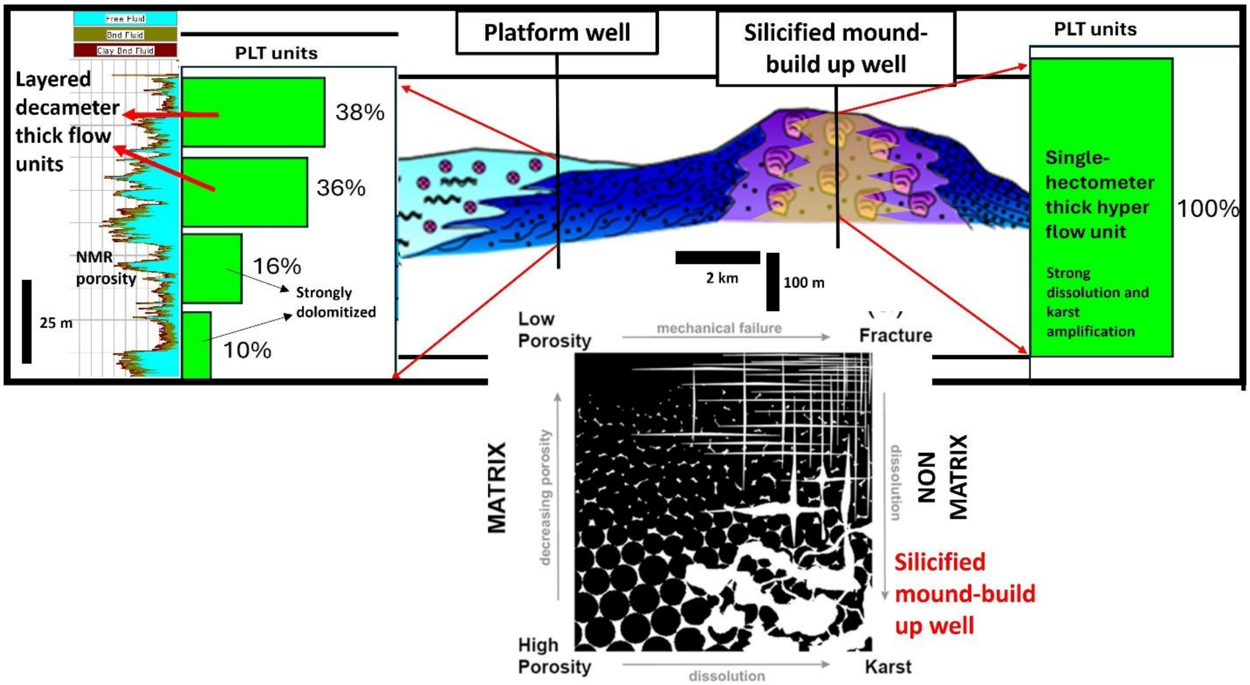

Figure 4.3 – Illustration of production logging tool (PLT) flow units' distribution according to depositional and diagenetic facies in the Mero field. Wells drilled in the platform exhibit layered flow units related to strongly dolomitized intercalated grainstones, algal mats, and spherulites (low depositional energy). The original depositional porosity is obliterated by superimposed episodes of diagenetic events (dolomitization and silicification). In the build ups and mounds wells (stromatolitic shrubstones) the dolomite cement was almost totally dissolved by silica rich hydrothermal fluids due to very high depositional porosity (interelement), and karst porosity that concentrated corrosive fluids circulation, the overall result is a net permeability gain, sometimes permeability excess (modified from Mimoun and Fernandez-Ibanez, 2023 and Maas et al., 2024).

Dataset

Our dataset consists of public domain well dynamic and seismic data from known producing fields of the Brazilian pre-salt province (Figure 4.1). As the training dataset (source), we used drill stem data (DST) from the Barra Velha (BVE) reservoir of the Mero oilfield, one of the main producing fields in the Santos Basin with test data for more than the thirty wells we used. We also have oil production history at the BVE reservoir from nine of those 30 wells. It is covered by 3D marine narrow azimuth seismic data from where we computed multiple post-stack attributes (complex trace, geometric, and voxel-based texture). We focused on seismic attributes because of their high availability during all phases of exploration and production (E&P) projects, rather than seismic inversion whose reliable results are only available in mature areas (with plenty of well information). As blind test datasets, we used DST and seismic attributes from Bacalhau and Lapa fields, less advanced production projects in the same basin, and similar geology, but different seismic surveys. Remember, the main objective here is to transfer the ML from the Mero field (source area) to other areas with different maturity and data availability (target areas).

In the Mero field, we used DST and production-injection test information from 30 wells (nine producers connected to a floating production storage and overload unit, FPSO). The well dynamic data are summarized in Figure 4.4, which shows the relationship between measured flow capacity (FC) and productivity index (PI) from the BVE reservoir (Figure 4.4a). Further details of this compilation can be found in Maas et al. (2024). The most important point is that, despite being based on transient flows, the DST flow capacity is an excellent proxy for reservoir permeability at seismic scale (R^2 = 0.90, Figure 4.4a) and daily oil production (R^2 = 0.83, Figure 4.4b) and the most prolific reservoirs are the silicified mounds and build ups which can produce up to 50,000 barrels per day (bpd) with a single vertical well. The average production at non-mound reservoirs is 10,000 bpd. The capacity of most FPSOs employed in Santos Basin deep waters is 180,000 bpd. Therefore, for a single FPSO full capacity fulfillment, only four wells drilled in hyper productive mounds are enough. On the other hand, dozens of non-mound producer wells are required. So, the earlier the hyper productive mounds are identified, the more the exploration and appraisal phases can be optimized. Maas et al. (2023) and Maas et al. (2024) used DST data to train and validate multiple ML models using seismic multi-attributes to predict flow capacity and seismic facies clustering. Herein, we used eight blind test validation wells (four at Bacalhau and four at Lapa field) for the evaluation of flow capacity ML-derived models using the training with Mero field DST data. Our seismic dataset is composed of three different seismic surveys (Fig 4.1) acquired at different times with different processing techniques that led to totally different amplitude values (Table 4.1). They were used as the input data for seismic multi-attribute derivation for ML training and prediction. The attributes used in this study are the Hilbert transform of input amplitude (HTAM), average frequency (AVF), sweetness (SWEET), the most positive curvature (K1), the most negative curvature (K2), energy ratio similarity (ERS), GLCM energy (GLEN), GLCM entropy (GLET), and GLCM homogeneity (GLEH). The theory behind these attributes can be found in Chopra and Marfurt (2007), and the reasoning for choosing them in Maas et al. (2023,2024).

Parameter	Mero RTM Acquisition – 2012 Processing - 2016	Bacalhau KPSDM Acquisition- 2002 Processing - 2014	Lapa KPSDM Acquisition – 2011 Processing - 2014
Bin size	25 x 25 m	25 x 25 m	12.5 x 12.5 m
Sample interval	5 m	5 m	5 m
Azimuth	360	270	300
Maximum offset	8,100 m	6,200 m	8,000 m
Amplitude range	-20,000 - +20,000	-40,000 - +40,000	-0.6 - +0.6

Table 4.1- Main parameters of the three different seismic surveys used in this study. Due to technical limitations of early 2000's marine seismic acquisition, seismic data over the Bacalhau field area was acquired with shorter cable length (6200 m) when compared to Lapa and Mero fields (>8000 m), for this reason, in the Bacalhau field, amplitude versus offset (AVO) data is compromised and, so, we did not use AVO attributes in this study. Acronyms: RTM (reverse time migration), KPSDM (Kirchhoff pre-stack depth migration).

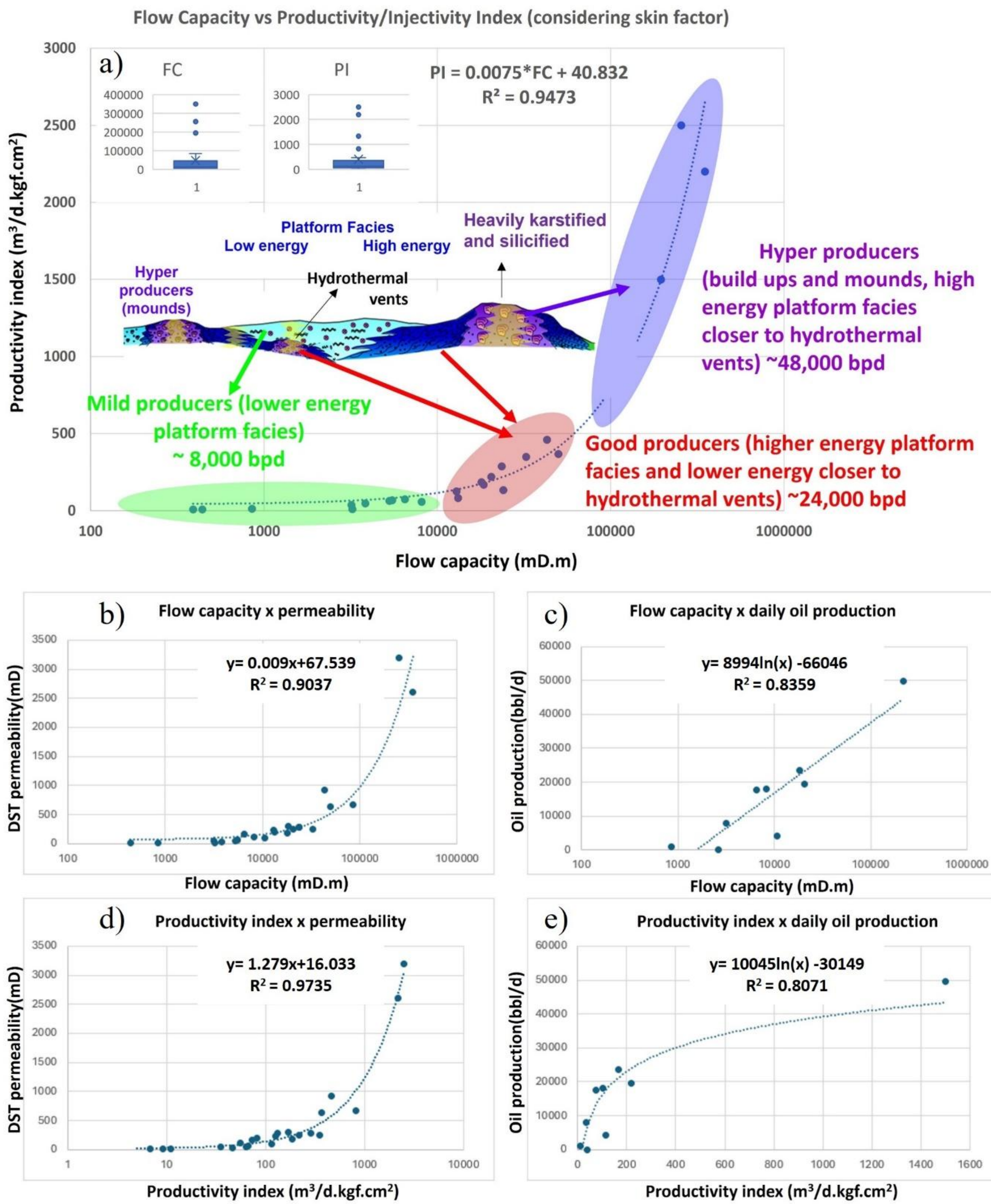

Figure 4. 4 – a) DST flow capacity, productivity index and production data from the Mero field. The BVE reservoirs can be classified into three major groups: hyper producers silicified build ups and mounds that produce up to 48,000 bpd, good producers form high energy platforms that produce around 24,000 bpd, and mild to low productivity lower energy platform facies producing 8,000 bpd on average (Maas et al, 2024). Second row: scatter plots of (b) flow capacity versus permeability (R^2=0.90) and (c) flow capacity versus daily production (R^2=0.83). Third row: scatter plots of (d) productivity index versus permeability (R^2=0.97) and (e) flow capacity versus daily production (R^2=0.80). Despite being acquired through short

periods of either production or injections (transient flows), the flow capacity is an excellent proxy for reservoir permeability prediction at DST and seismic scale, as well as for production forecasting.

Method

From previous research, we have both lithology and productivity ML derived models for the BVE reservoir from the Mero Field (Maas et al., 2023, and Maas et al., 2024). These models are validated with blind test wells, and we want to transfer the ML to other seismic datasets in two other oilfields hundreds of kilometers distant from Mero Field to predict the same reservoir parameters considering the BVE reservoir is roughly similar across the different areas in the same basin (Santos Basin). At these two other areas, there are well dynamic data that was used to evaluate our predictive models. As will be explained later, our transfer learning approach relies in the fact that geology does not vary too much across the different areas (Mero, Bacalhau, and Lapa). Indeed, there are some local geological discrepancies that were handled via feature engineering and hyperparameter tuning. Therefore, the main issue for transfer learning is the different seismic data that were acquired and processed in different times using different technologies and parameters. The way we used to handle this drawback is to focus on the Barra Velha reservoir by using the top and bottom surfaces interpreted from seismic data and normalize all seismic attributes in the same range of those from the Mero field seismic survey. Our work hypothesis is that this approach is enough to test the machine learning algorithms trained with Mero field seismic attributes and well data in other areas with similar geology covered by very different seismic surveys.

From the ML algorithms tested at the Mero field in our previous research, we selected the best ones (SOM and RFR), whose training hyperparameters were transferred to Bacalhau and Lapa datasets for supervised regression of flow capacity and unsupervised clustering for facies mapping. Our method is summarized in Figure 4.5. The starting points of our analysis are the SOM clustering and random forest regression of flow capacity models that fit the Mero field training dataset. The SOM clustering model was built without any well training, just multi-attributes hyperdimensional clustering and was validated using Mero field DST data (Maas et al., 2023). The random forest regression model was built using DST information from ten training wells, ten seismic attributes around perforated zones, and validated using grid search cross validation (GSCV) within the training dataset and then blind test validation using 20 more wells that were not used for any training. The blind test performance of this model at the Mero field is 80 % (16 out of 20 correct predictions, Maas et al., 2024).

Transfer learning basic aspects

The first articles on transfer learning were published in the late 70's to 90's of last century (Bozinovski and Fulgosi, 1976; Bozinovski et al., 1977; Bozinovski,1985; Caruana, 1997 and Thrun and Pratt, 1998). Transfer learning is nothing but to training and validating an ML algorithm within a specific dataset and use this training to make predictions in a similar dataset that may differ from the training one in many aspects, such as feature space and distribution (Bozinovski, 2020). Despite these differences between the source (training) and target (blind test validation) datasets, the pre-trained and validated algorithm's features and hyperparameters are simply used to make predictions in a new target dataset, skipping adding new data points (data augmentation) to the dataset and retraining it, which can be time and money consuming. The "machine learned" hyperparameters and related parameters are transferred from a task to a new similar one.

According to Pan and Yang (2010), one big assumption is that the source and target datasets must be somehow alike to ensure the success of TL. Huang and Tian (2024) describe TL in the following equation:

$$\begin{cases} D = \{X, P(X)\} \\ T = \{y, f(.)\} \end{cases}$$

(1)

Where D and T are domain and task, respectively. The domain D is where the learning will be transferred from, and T is the task to which it will be transferred. X and y are feature and target spaces respectively; P(X) is the marginal probability distribution, and f(.) is the prediction function, the ML learning algorithm itself. According to them, this scheme is enough to allow the learning acquired from the source domain to the target domain, considering the similarity conditions stated by Pan and Yang (2010) are satisfied. In our case, we selected from the works of Maas et al. (2023) and Maas et al (2024) their best algorithms used for unsupervised clustering for seismic facies classification and random forest regression (supervised learning) trained with Mero field well data and seismic multi-attributes. The Mero field is the source dataset and the Bacalhau and Lapa fields are the target datasets. For unsupervised clustering, we used the same input seismic attributes and training parameters for SOM clustering done in the Mero dataset and applied them to the Bacalhau and Lapa ones. Likewise, we used the same input seismic attributes and random forest regression hyperparameters which led to the best validation performance in the Mero dataset and applied to Bacalhau and Lapa to predict the flow capacity at BVE reservoirs. For transfer learning evaluation, we used DST information from four wells in each target dataset as blind tests totaling eight validation wells.

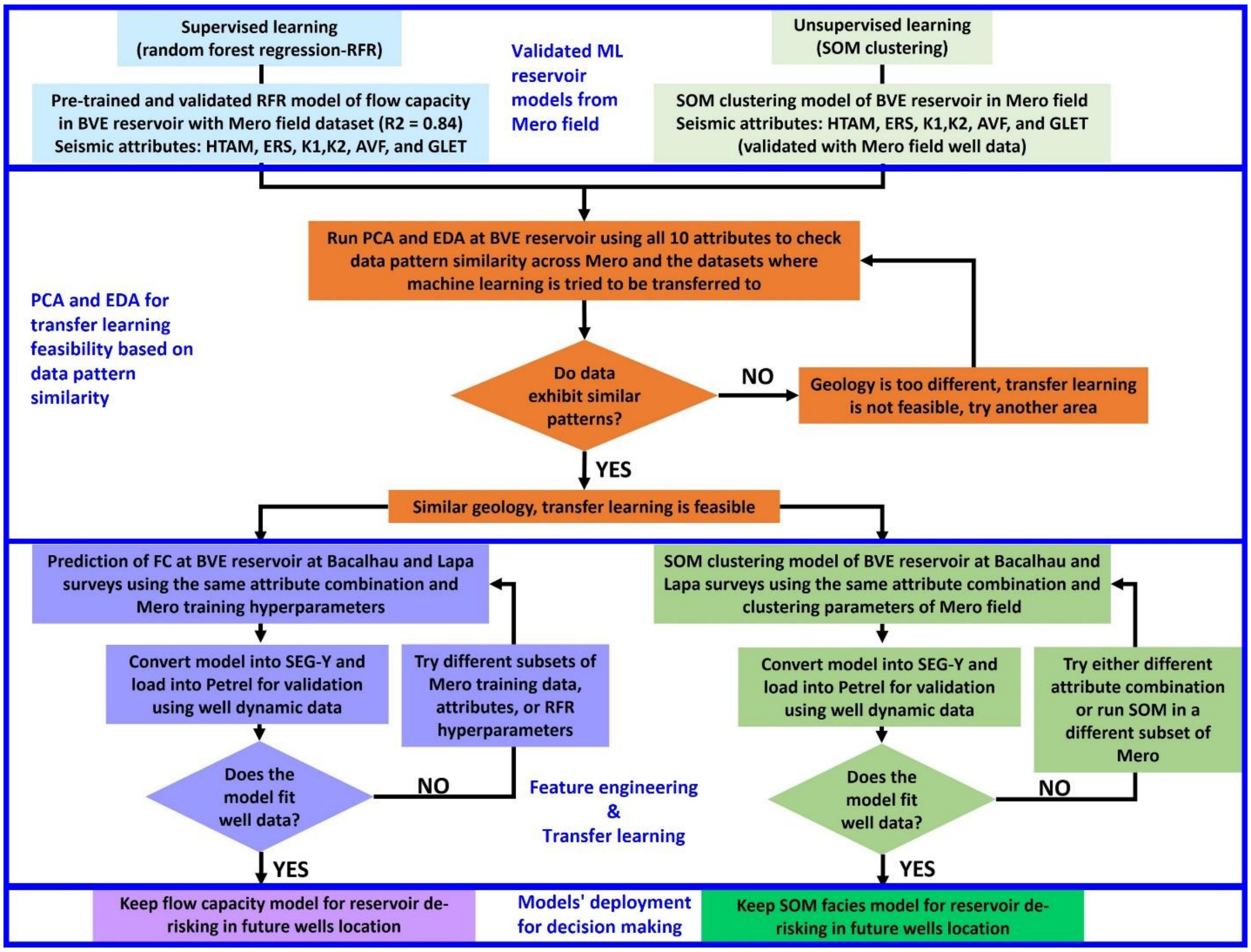

Figure 4.5- General workflow of this research.

Transfer learning feasibility studies

The transfer learning feasibility studies started with principal component analysis (PCA) and exploratory data analysis (EDA) of the attributes common to the three datasets within the BVE interval. Because of the short cable of Bacalhau seismic acquisition, the AVO information was compromised, so AVO attributes were not computed for the Lapa and Bacalhau datasets. This situation is quite common in exploratory areas, where reliable AVO and inversion information is not always available.

Principal component analysis (PCA)

To check the data similarity across the three different seismic datasets, we ran PCA analysis (Pearson, 1901; Hotelling, 1933; Jolliffe and Cadima, 2016, and Ha et al., 2021) with nine attributes common to all datasets (HTAM, SWEET, AVF, K1, K2, ERS, GLEN, GLET, and GLEH). Because of its lower sensitivity to outliers, we normalized the attributes using standard normalization (mean=0, standard deviation=1), and then ran PCA between the top and base of the BVE reservoir interval by using the top and base of the BVE reservoir interpreted surfaces as model outlines. Figure 4.6 shows the attributes' contribution to the variability of the first eigenvector of the covariance matrix within the BVE reservoir. In the three different datasets, the attributes' importance is very similar, furthermore, there is a consistent pattern of PCA importance variation across the datasets. It is an indication that, due to the geological similarity at the BVE reservoir, the seismic attributes depict the main lithological variations common to the three different seismic surveys.

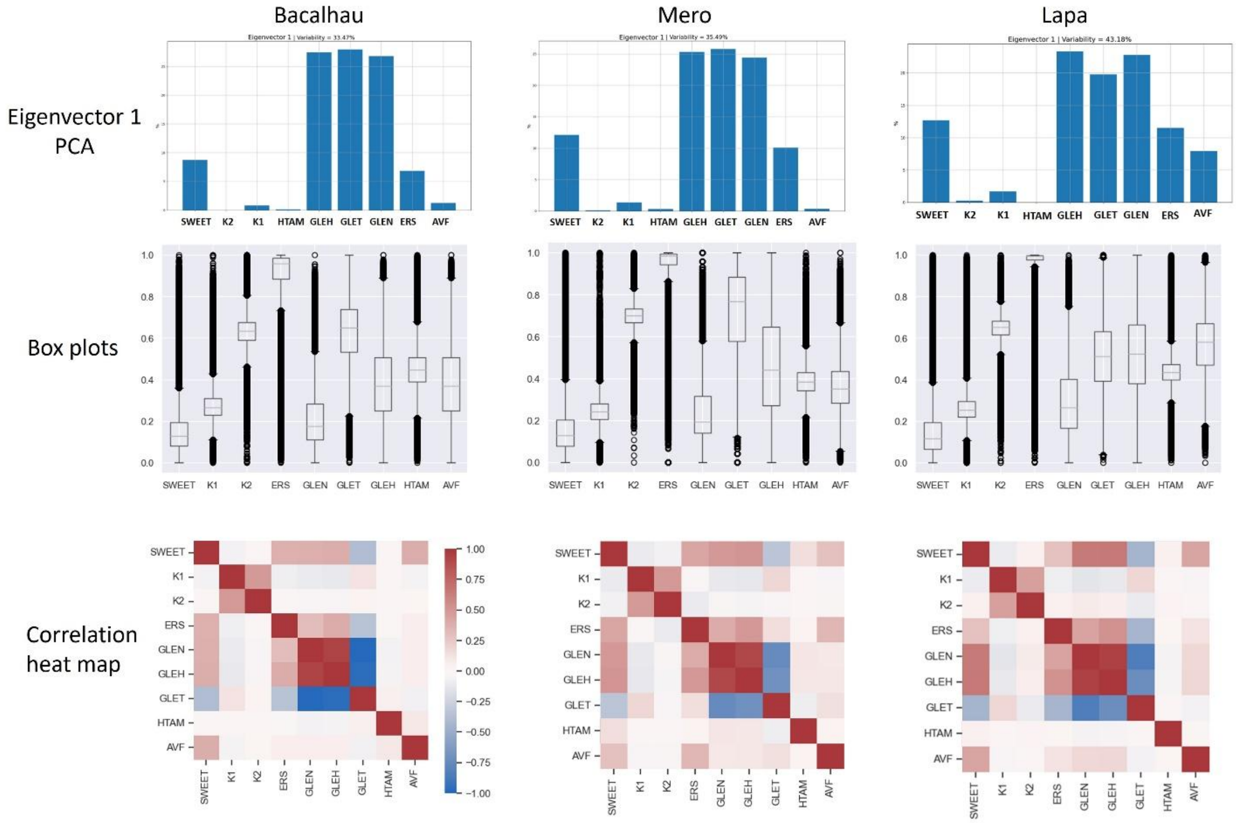

Figure 4.6- First row: first eigenvector PCA of the nine seismic attributes common to the three different surveys at the BVE reservoir level. The PCA importance of the attributes for eigenvector's variability is persistent across the different surveys. Second row: box plots of the normalized attributes within the BVE zone. The general distribution and skewness are also persistent in the three different seismic surveys. Third row: the correlation heat map pattern is similar across the different seismic surveys. All this PCA and statistical analysis indicate high geological similarity between Mero field (source) and Bacalhau and Lapa fields (target datasets).

Exploratory data analysis (EDA)

The same nine attributes used in PCA were min-max (0 to 1) normalized for EDA using the Python library Seaborn freeware. Inside the BVE zone, in all three different seismic surveys, we computed the correlation matrix and the boxplots of the normalized attributes which are shown in Figure 4.6. We can observe a recurrent pattern similarity across the different seismic surveys, as the data skewness observed at Mero field attributes is somehow present in the Bacalhau and Lapa field ones (boxplots of Figure 4.6) and the correlation heatmap pattern is also similar across the 3 different surveys. This is a good indication that transfer learning is feasible from the source area (Mero field) to the target areas (Bacalhau and Lapa Fields) because of the overall geological similarity between them. This can be observed in Figures 4.7 and 4.8 which show respectively the gross depositional environment (GDE) of the BVE reservoirs and seismic-based geological sections across the three oil fields. In Figure 4.7, we can see the strong similarity between the Mero and Bacalhau fields in terms of facies distribution. Both Mero and Bacalhau carbonate platforms exhibit important build ups and mounds aligned with basement inherited faults that also functioned as conducts for strong hydrothermal silicification that enhanced their pre-existing karst porosity. At both Mero and Bacalhau fields, there is a tectonic deformation (syn-salt deposition oblique extensional fault reactivation) affecting the BVE reservoirs which can be seen in terms of fault density and silicification intensity. At the platform facies of the Mero Field, the carbonates are dolomitized, the silicification is not pervasive, and it is confined to the reactivated faults, where the dolomite cement is dissolved giving place to silica precipitation. In general, this process is more intense in the Bacalhau field area (Figures 4.8b and 4.8c). The Lapa field area is a less tectonically deformed area, where low energy flat platform-related facies are dominant with rare and sparse mounds confined to the Well L2 area (Figure 4.7), the fault density is low and most of the BVE reservoirs are strongly dolomitized, resulting in low productivity. The higher flow capacity values are only observed at wells close to major faults where the dissolution related to silica-rich hydrothermal fluids is more intense like in the Mero field (Figure 4.8a).

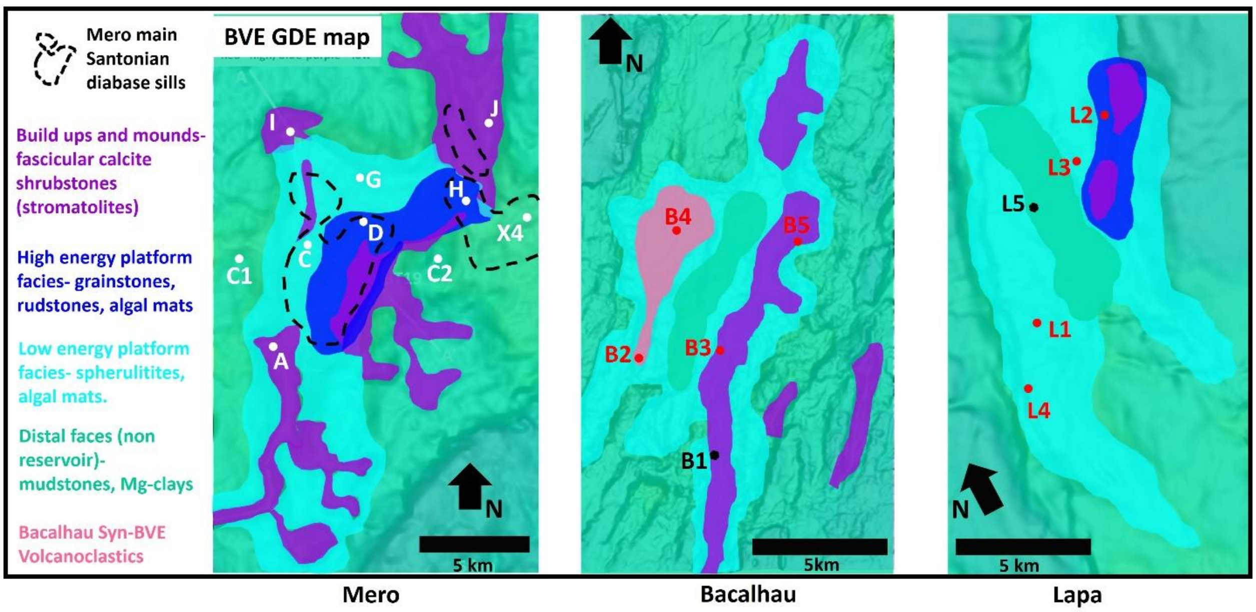

Figure 4.7 – Gross depositional environment (GDE) map of BVE reservoirs at Mero, Bacalhau and Lapa fields. All three areas have the same type of lithologies with different volumetric proportions. At the Mero and Bacalhau fields, the build ups and mounds are expressive. In the Lapa field, the most representative facies are the low depositional energy reservoirs. In the Bacalhau field, there are syn-BVE volcanoclastic mounds that do not occur in other areas. In the Mero field, we have Santonian age sill intrusions in BVE reservoirs that were not yet found in Bacalhau and Lapa fields.

Although we have shown numerical and geological similarities among the three datasets, there are important differences that impact the reservoir productivity. At the Mero field, there is Santonian magmatism in the form of diabase sills in different stratigraphic levels, including the BVE reservoir, where they act as both flow and drilling barriers. These Santonian age intrusive bodies were not yet intercepted in the Bacalhau and Lapa fields

(Figures 4.7 and 4.8). At the Bacalhau field, there is a syn-BVE (late Aptian) volcanoclastic mound (mostly ignimbrites) that is non-reservoir but has seismic amplitude signatures remarkably like the prolific carbonate mounds. This volcanoclastic mound can be interpreted as the final stage of initially tholeiitic basaltic volcanism (Parati basalts, Moreira et al., 2007) that evolved to more acidic volcanism during the deposition of lower BVE carbonates in this sector of the Santos basin (Figure 4.8c). This is an important pitfall that still pursues the operators in other areas of the Brazilian pre-salt province and worldwide. Reducing or eliminating the ambiguity of seismic prediction of productive reservoirs in this kind of challenging geological setting is critical for E&P projects optimization. Additionally, the Lapa field differs from the Bacalhau and Mero fields because of the scarcity of hyper productive mounds and build ups. It is more like the low energy platform of the Mero field, characterized by layered, strongly dolomitized, and, thus, low productivity carbonates (Figures 4.7 and 4.8). These similarities and discrepancies forced us to adopt some procedures regarding the training dataset and input seismic attributes configurations to optimize the transfer learning efficiency when using Mero training hyperparameters for productivity prediction at the Bacalhau and Lapa datasets. These procedures involving training datasets and input attributes (features) manipulation are known in the literature as feature engineering (Rawat and Khemchandani, 2019).

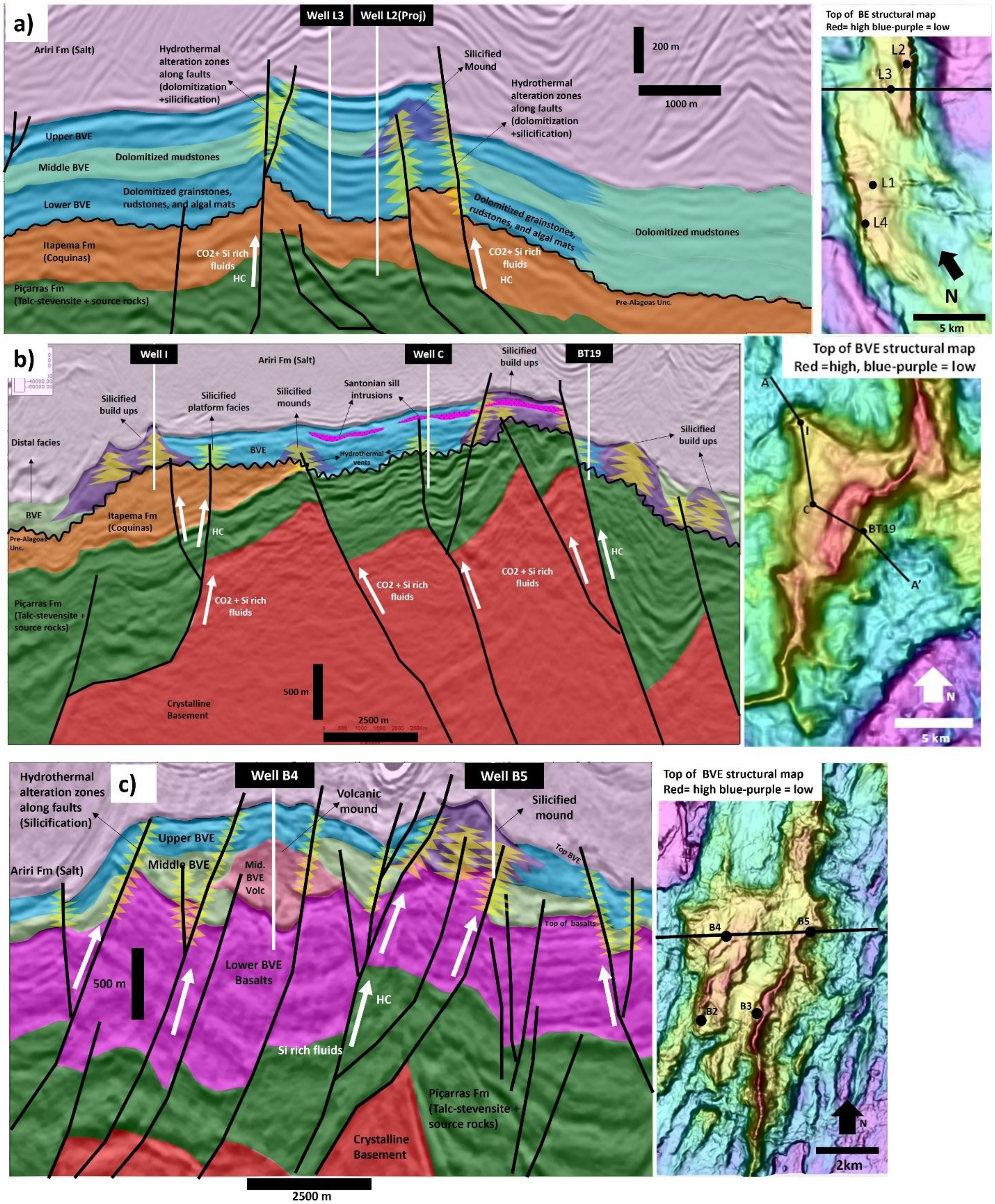

Figure 4.8- Seismic and well interpretation based geological sections of Lapa field (a), Mero field (b), and Bacalhau field (c). From Lapa to Bacalhau, there is an increasing fault density that affects the BVE reservoir. As a result, the build ups and mounds presence as well as their hydrothermal silicification is more prominent in Mero and Bacalhau fields. This reflects the higher productivity observed in these last two areas.

Feature engineering for transfer learning optimization

Figure 4.9 shows an illustration of the Mero field training dataset for random forest regression. Detailed information on its building can be found in Maas et al. (2024). As training dataset, we used flow capacity (FC), productivity index (PI), and seismic attributes data along the DST perforated zones in the wells. Due to their similarity to Lapa field wells, (lower energy platform dolomitized facies predominance) we separated the mild and low productivity wells (blue and green in Figure 4.9) into a training subset (Mero low) and retrained our algorithms for FC prediction at Lapa field. The red cube is a subset of the Mero seismic survey where the retrained model was validated using ten blind test wells (white wells in Figure 4.9) because of their high similarity to Lapa field. For predictions in the Bacalhau field, we used the original Mero training dataset with all wells (hyper, mild and low productivity- The Mero high subset) because of their high similarity. Whenever available, production logging tool data (PLT) were also used. PLT data show the flow units that contribute most to observed production rates during DST, and consequently the higher FC units (Figure 4.10). In the case of well C, almost the entire carbonate section of BVE was perforated and has PLT data. The only non-tested zone was the diabase sill intrusion which has no porosity. The training dataset was completed with zero values of FC in this interval to counterbalance the bias due to only non-zero measured FC values. It also increases the variability in the training dataset, which can improve the machine learning prediction of flow capacity. Maas et al (2024) used this dataset as training for multiple supervised regression models. The best result (85 % blind test performance) was achieved with random forest regression of FC using seismic attributes HTAM, AVF, K1, K2, ERS and GLET as predictors. Maas et al. (2023) used these attributes for unsupervised clustering for seismic facies classification in the BVE reservoirs of the Mero field. The best result was obtained with SOM clustering, which was able to separate the hyperproducer mounds and build ups from the platform and distal facies, and the diabase sill intrusions in different SOM clusters, all validated with well information which was not used for any training in this case (unsupervised learning). These two models are shown in Figure 4.11, and they are the starting point for our transfer learning strategy. Because of data and general geology similarities between the Mero and Bacalhau fields, we ran RFR prediction of flow capacity and SOM clustering at the Bacalhau field by simply using the same attributes and ML parameters that were optimized in the Mero field. We got interesting results that will be detailed in the next sections. However, when applied to the Lapa field, it led to unrealistic results that did not fit the existing well data (mostly overpredictions). This is due to the geological differences between the Mero and Lapa fields primarily related to GDE distribution, as the Mero field has huge build ups and several minor mounds in its platform, whereas at the Lapa field, these mounds are small, sparse and confined to a specific compartment (Figure 4.7).

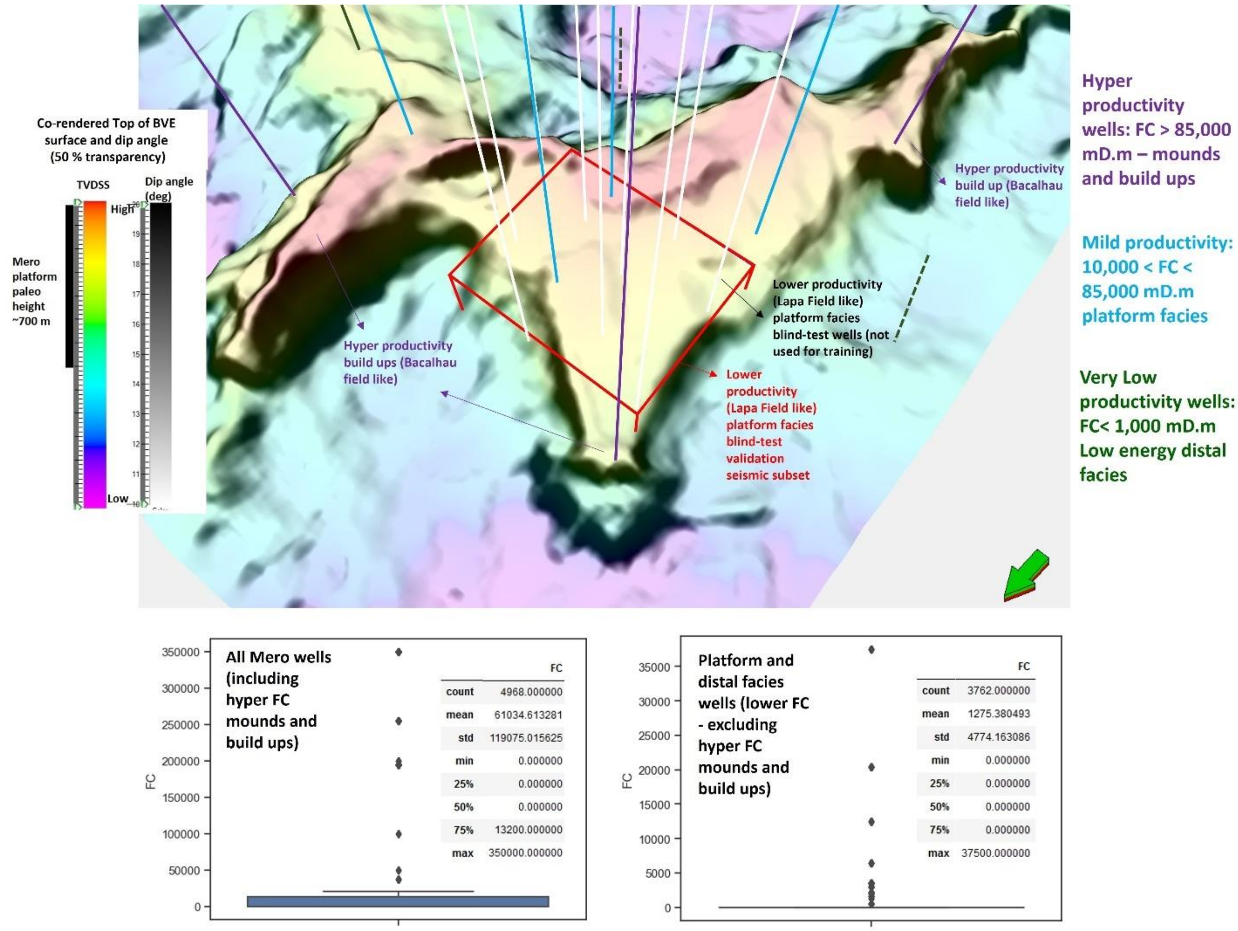

Figure 4.9 – 3D view Mero field carbonate platform. The surface is the top of BVE reservoir co-rendered with structural dip to enhance the faults (black shaded areas). Wells in purple are the highest productivity ones (silicified mounds and build ups), wells in blue are mild productive (dolomitized platform facies, locally enhanced near faults), and wells in green are the lowest productivity wells related to dolomitized distal facies. All these ten wells were used as training dataset where their DST flow capacity values were used to train supervised regression algorithms (Maas et al., 2024). At the bottom of the figure, we have the box plots of FC distribution in the Mero high (0 to 350,000 mD.m) and Mero low training subsets (0 to 37,500 mD.m).

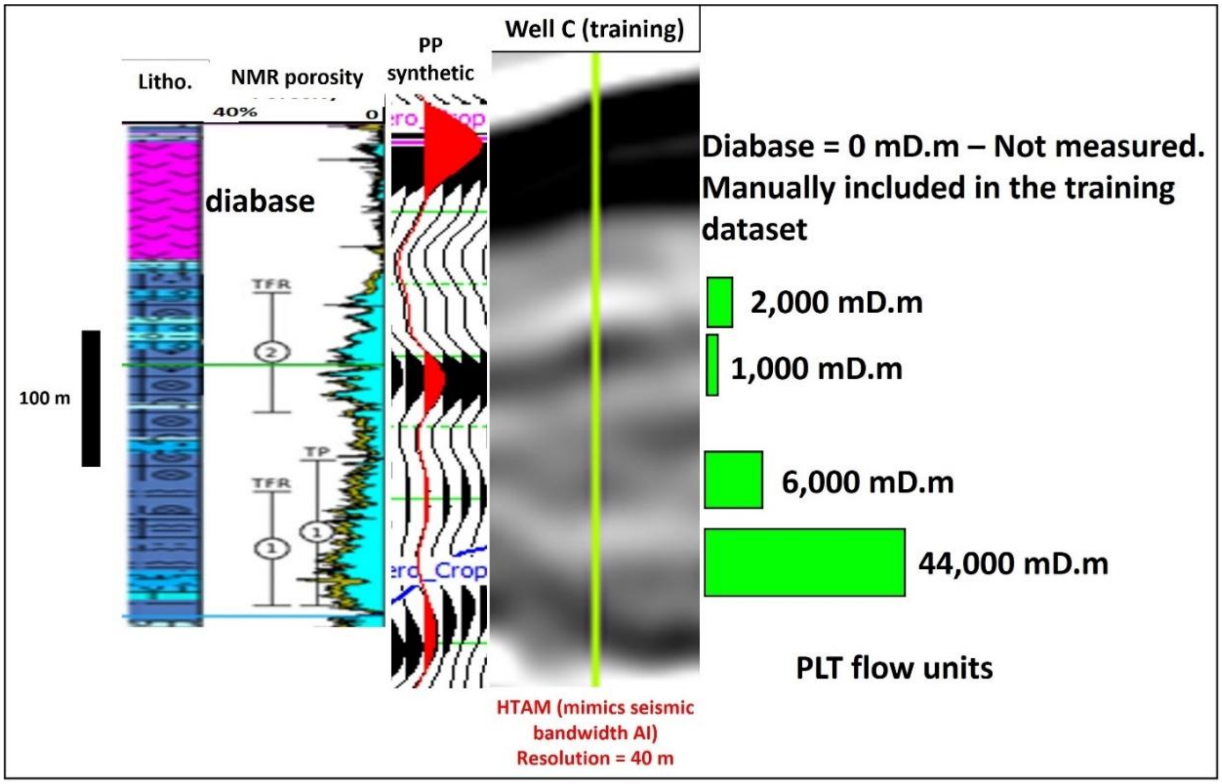

Figure 4.10- Composite log (left), PP synthetic seismogram and HTAM attribute vertical section at one of the Mero training wells location (Well C), which was drilled in the low energy platform. We used DST and PLT flow units from perforations at BVE reservoirs and seismic attributes along with the perforations. As can be seen, the PLT flow units' thicknesses are compatible with seismic resolution, which is 40 m. Because of its zero porosity and proximity to the other perforated zones, we inserted in our training dataset non-measured zero values of FC to counterbalance the biasing of the training dataset which is composed of mostly non-zero values. This can also help our supervised learning algorithms to recognize other types of igneous bodies as nonproductive intervals.

The Mero and Lapa fields also differ in terms of tectonic deformation as the Mero field has a higher fault density than Lapa (Figure 4.8). As a result, the hydrothermal alteration along faults is also more intense in the Mero field. The only aspects that Mero and Lapa field have in common are the lower energy and strongly dolomitized platform facies. To address this issue, we selected a subset of the original Mero field training dataset which sampled only platform and distal facies (strongly dolomitized, lower FC values, herein named Mero low subset), excluded the hyper producing wells (build ups and mounds, highest FC values, herein named Mero high subset). By doing this, we reduced the size of the training dataset from 4,968 to 3,762 datapoints, not a big training data loss. Next, we re-ran random forest regression in this lower productivity wells subset (Mero low) and got similar best hyperparameters, but totally different random forest and Shapley additive explanation (SHAP) attributes importance when compared to the original dataset with all wells (Mero high, Figure 4.12). SHAP is a ML models explanation technique which uses the cooperative game theory to explain how each attribute individually contributes to the modeled output (Lundberg and Lee, 2017; Lundberg et al.,2018, 2020; Molnar, 2021 and Lubo-Robles, 2022). In the Mero low subset, (Figure 4.12b) the best RFR model of FC that fits the data has an attributes importance different from the Mero high training subset (with hyper flow wells, Figure 4.12a). In the low energy platform facies (and low productivity), which is more like the Lapa field, the main reservoir type is the layered flow units described in Figure 4.3. It is characterized by strongly dolomitized interbedded in situ microbial mats, grainstones, and spherulites. As can be observed in the random forest and SHAP importance (Figure 4.12b), the most important attributes are now AVF, SWEET, HTLR and HTAM. Because of the absence of hyper productive mounds and build ups, and the dominance of low energy dolomitized layered reservoirs, the highest FC values are related to lower impedance and frequence attenuated reservoirs closer to the faults. The lowest FC values in both cases are related to strongly dolomitized low energy platform facies (higher SWEET and HTAM values).

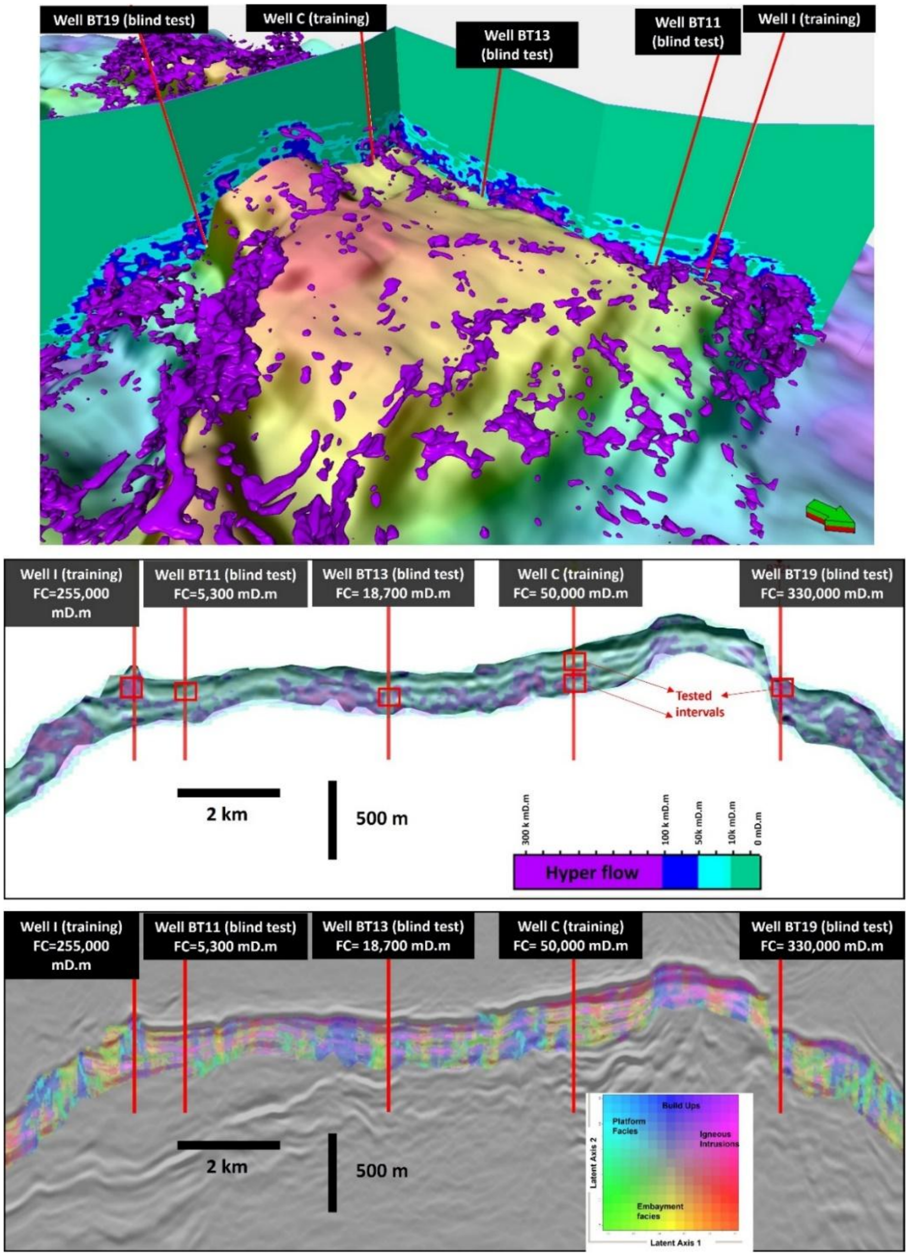

Figure 4.11- Top: 3D view of the predictive model of flow capacity (FC) (using random forest regression- RFR) at the Mero field showing some of the training and blind test wells (Maas et al., 2024). The purple geobodies are the predicted hyper flow zones mostly related to heavily silicified build ups and minor mounds. Middle: same arbitrary seismic section of HTAM attribute with RFR predicted FC. Bottom: same arbitrary section showing SOM clustering unsupervised model and its interpretation. Note the coincidence of high FC zones with the clusters related to build ups and mounds and the low FC zones with clusters related to sill intrusions. For both RFR and SOM clustering, the input seismic attributes used were HTAM, AVF, K1, K2, ERS and GLET.

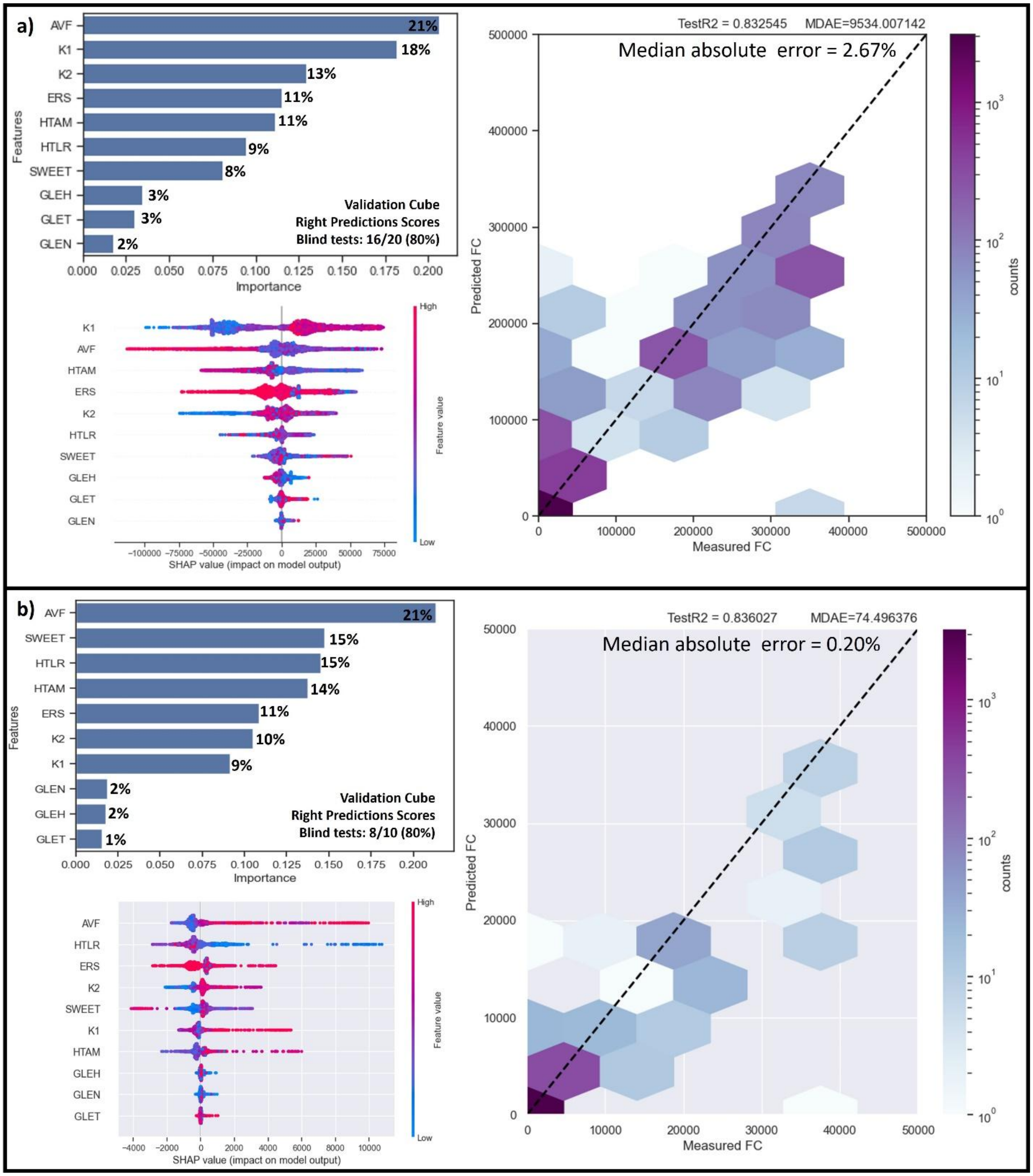

Figure 4.12 – a) Random Forest regression model of FC using all 10 attributes in the Mero high training dataset (with hyper flow units). Both random forest and SHAP attributes' importance show that due to the presence of hyper productive build ups and mounds, the most important attributes are AVF, K1, K2 and ERS due to their convex shape, and frequency attenuation due to intense hydrothermal karstification. The hexagonal bins plot of predicted versus measured FC shows a good fit between the predictive model and the training data (median absolute error = 2.67% and R^2=0.83). b) the same as A but the training dataset is the low productivity Mero subset. In this case both SHAP and RFR attributes importances changed when compared to the Mero high dataset, due to the dominantly layered nature of the reservoir affected by strong dolomitization, leading to lower productivity.

Closer to the faults, there is an incipient hydrothermal corrosion that was not pervasive due to lower porosity related to depositional facies and superimposed burial dolomitization. The mounds are scarce, smaller, and restricted to faults. For this reason, the attributes, K1, K2, and ERS (geometric) have much less importance than the complex trace ones (HTAM, SWEET, and HTLR) for FC prediction when compared to the prediction using data from Mero high training dataset (Figure 4.12a). In this case, the most important attributes were AVF, K1, K2, and ERS due to the convex shape of hyper productive mounds, and their intense hydrothermal karstification that causes frequency attenuation. For this reason, we fit a different RFR model of FC using SWEET, HTAM, ERS, K1, and GLEN that needed more, and deeper (500 regression trees, max depth=12) regression trees compared to the 400 trees with a max depth of 10 that fits the Mero high subset. The GSCV R^2 of the model that fit the Mero low subset is 0.83 and its blind test performance in the validation seismic subset described in Figure 4.9 (red cube) is 80% (8 out of 10) correct predictions of FC in the blind test wells positions (Figure 4.12b). We also ran SOM clustering in this seismic subset using the same attributes (Figure 4.13) to see the different facies patterns. Next, at the transfer learning stage, we ran RFR of FC and SOM clustering in the Lapa field dataset using the Mero low training with this new attribute's combination. As mentioned before; to honor the premise of data similarity (Pan and Yang, 2010), we had to perform the above-described human knowledge-driven feature engineering to make accurate predictions that will be shown in the next section.

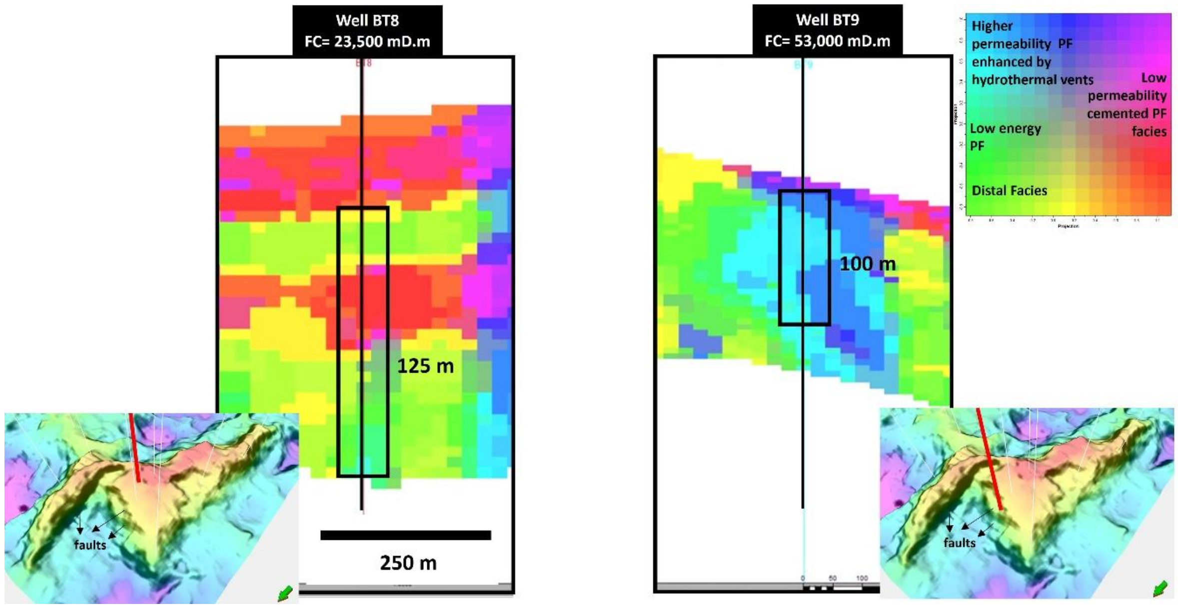

Figure 4.13 – Mero Field: SOM clustering model using SWEET, HTAM, ERS, K1 and GLEN seismic attributes as inputs. At the blind test well BT8 position, we have the cumulative FC of the tested interval (black rectangle, 125 m thick DST interval, FC=23,500 mD.m). As this well was drilled far away from the major faults, the BVE reservoirs are strongly dolomitized low energy platform facies (higher impedance, red-pinkish clusters), intercalated with less dolomitized layers (lower impedance, green-yellowish clusters). The BT9 well was drilled close to a fault zone, here the reservoirs are less dolomitized with more cement dissolution due to hydrothermal activity along faults. The FC values in the DST interval (black rectangle, 100 m thick) are the highest ones in the low platform energy (53,000 mD.m). The SOM clustering depicted these hydrothermally enhanced reservoirs as subvertical bluish clusters (fault controlled).

Through this systematic approach of feature engineering and model optimization, we developed two distinct workflows for transfer learning: one using the complete Mero training dataset for application to the geologically similar Bacalhau field (Mero high), and another using a subset of lower-productivity wells for the Lapa field predictions (Mero low). These workflows may be adopted depending on the interpreter's judgement (human seismic interpreters jobs matter!) on the geological similarity between the source and target areas. The effectiveness of these workflows was evaluated using blind test wells in each field, with success measured by the ability to predict both flow capacity values and facies distributions. In the following section, we present the results of applying these workflows, first to the Bacalhau field where we achieved 85% prediction accuracy, followed by the Lapa field where we obtained 75% accuracy in flow capacity predictions.

Results and Discussion

In this section, we show the TL predictive models in the Bacalhau and Lapa fields. We focus on SOM clustering (unsupervised learning) and RFR of flow capacity (supervised learning) models of the BVE reservoirs in those areas covered by different seismic surveys. For SOM clustering models, at Bacalhau field, we used the same training parameters and attributes combinations used at the whole seismic survey of the Mero field (Maas et al., 2023). For the SOM clustering at the Lapa field, we used the same parameters used in the seismic subset at the lower energy platform of the Mero field (red cube in Figure 4.9, SOM model in Figure 4.13). For RFR of flow capacity, at the Bacalhau field, we used the same attributes combination and hyperparameters that best fit the Mero high training subset (Figure 4.12a, Table 4.2). At the Lapa field, we used the same attributes and RFR hyperparameters that best fit the Mero low training subset (Figure 4.12b, Table 4.2). We used four wells in the Bacalhau Field (B2, B3, B4, and B5) and four wells at the Lapa field (L1, L2, L3, and L4) as blind tests for both SOM clustering and RFR of flow capacity models. For SOM clustering model validation, we checked in each well position at the BVE interval and DST perforations (where available) whether the clusters of the SOM model correspond to the lithology and /or PLT flow zone (where available). For our RFR models of flow capacity validation, we checked in each perforated zone to see if the measured flow capacity value lies in the predicted flow capacity interval. For instance, if we got a predicted flow capacity value of 5,000 mD.m and the measured flow capacity is 1,000 mD.m, we consider it as a correct prediction as both measured and predicted flow capacity values lie in the same interval (0-

10,000 mD.m) that corresponds to very low productivity reservoirs in the Mero training dataset. The other flow capacity intervals are: 10,000-50,000 mD.m (medium to high productivity), 50,000-100,000 mD.m (high productivity), and > 100,000 mD.m (hyper productivity).

Our research results are summarized in Table 4.2, where we have the main parameters of each predictive model calculated for the Bacalhau and Lapa fields (target areas) using training parameters obtained with Mero field data (source area). We transferred the learning from the source to the target areas by using different training strategies considering geological similarity.

Random forest regression FC model	Input attributes	Random forest hyperparameters	# data points	Mean FC (mD.m)	Prediction performance
Mero training with hyper FC	HTAM, ERS, K1, K2, AVF, GLET	Max depth = 10 # decision trees =400	4,968	61,034 (measured)	83% (GSCV R^2)
Mero training subset without hyper FC	HTAM, SWEET, K1, ERS, GLEN	Max depth = 12 # decision trees =500	3,762	1,275 (measured)	83% (GSCV R^2)
Mero blind test validation with hyper FC	HTAM, ERS, K1, K2, AVF, GLET	Max depth = 10 # decision trees =400	2,519,916	52,607 (predicted)	80 % (16 out of 20 right predictions)
Mero blind test validation subset without hyper FC	HTAM, SWEET, K1, ERS, GLEN	Max depth = 12 # decision trees =500	1,775,193	3,480 (predicted)	80 % (8 out of 10 right predictions)
Bacalhau blind test validation (from TL)	HTAM, ERS, K1, K2, AVF, GLET	Max depth = 10 # decision trees =400	2,518,916	30,664 (predicted)	87.5 % (3.5 out of 4 right predictions)
Lapa blind test validation (from TL)	HTAM, SWEET, K1, ERS, GLEN	Max depth = 12 # decision trees =500	6,186,735	187 (predicted)	75% (3 out of 4 right predictions)

Table 4.2- Results summary of random forest regression of fault capacity transfer learning from Mero to Bacalhau and Lapa fields. The two first rows are the models fitted to training subsets with (Mero high FC - Bacalhau like) and without (Mero low FC- Lapa like) hyper productive wells from Mero Field. The third and fourth rows are respectively the predictive models in the Mero high FC and Mero low FC (red cube in Figure 4.12). The fifth and sixth rows are the predictive models using transfer learning to Bacalhau and Lapa fields. The overall blind test performance in the three seismic surveys is around 80%, which is very close to the GSCV performances during training. This is a very good indication of the high generalization capacity of our predictive models trained with Mero field data.

Bacalhau field application

At wells B2 and B3, B4, and B5, our RFR model has made accurate predictions, totaling six out of seven right predictions (85% performance) of flow capacity at the perforated zones (Figure 4.14c, Table4. 2). All this by training without any feature engineering (Mero high training subset). As mentioned before, the Bacalhau field has high similarity with the Mero field in terms of the volumetric importance of the hyper producing build ups and mounds (FC > 100,000 mD.m) (Figures 4.7 and 4.8). In the same way, there is a strong hydrothermal influence along major fault segments, which is confirmed by higher FC values close to either faults or hydrothermal vents (Figure 4.8b and 4.8c). In the Bacalhau field, Wells B3 and B5 are the most productive ones (166,000 and 135,000 mD.m FC values respectively), they drilled heavily silicified build ups whose cave permeability was overall enhanced and only partially obliterated by cryptocrystalline and mega-quartz growth. Similar processes are observed in the Mero field. Therefore, we applied the random forest regression of flow capacity shown in Maas et al. (2024) using the same "learned" hyperparameters from the Mero field. Then, we built a predictive model of FC using seismic attributes HTAM, AVF, ERS, K1, K2, and GLET as predictors in the seismic dataset cropped around the Bacalhau field, where there are DST information (FC and other parameters) of 4 wells which were used as blind tests. We also ran SOM clustering with the same attributes and parameters from the Mero field (Maas et al., 2023).

Figure 4.14a shows a vertical section passing through Bacalhau Field blind test wells B3 and B4. The HTAM attribute overlain by RFR model of FC obtained using Mero training. Figure 4.14b shows the same as 4.14a but the overlay is an SOM clustering model using the same attributes and parameters as Mero Field (training dataset). In Figure 4.14a, one can see that in Well B4, the FC values at the perforated zone (blue rectangle) were overpredicted by this model (measured 5,700 mD.m versus predicted above 10,000 mD.m). However, at the volcanoclastic zone (ignimbrites), which was not perforated due to its low porosity, the predicted FC values are dominantly between zero and 1,000 mD.m (extremely low), as expected for this kind of lithology. At this well location in Figure 4.14b, despite not being classified within a specific cluster, the volcanoclastic mound does not have the same cluster signature as the hyper productive carbonate mounds and in terms of geobodies, they do not form continuous features (Figure 4.14c). These smaller mounds are geologically feasible to be present in volcanic edifices (atoll facies) but are volumetrically unimportant. In the perforated zone of Well B4, SOM clusters have depicted low energy dolomitized platform facies. Both SOM and RFR models have correctly predicted the hyper productive silicified build up occurrence and the FC values in the perforated zone at Well B5. For the above-mentioned reasons, in our RFR model assessment, we consider the Well B3 prediction to be half successful (bad in perforation but accurate in volcanoclastic) and the Well B4 prediction a success. As mentioned before, seismic discrimination between productive carbonate mounds from volcanic non-reservoirs with mound geometry is an old challenge that still pursues the industry. Our RFR flow capacity model was not trained with volcanoclastic rock well data (it was derived using supervised regression rather than a classification algorithm). It was trained with thousands of datapoints of very low flow capacity carbonate reservoirs and artificially included non-productive diabase sills (without DST perforations). This allows that, in terms of multi-attributes response, the algorithm to predict either tight or non-reservoirs of different nature as very low flow capacity zones. In the Bacalhau field, our TL models trained with Mero data could accurately predict the most important lithologies and flow units. Nevertheless, it is not a perfect tool (85 % blind test performance) as it overpredicted a low productivity zone at one of the perforations (Well B4). Even though, it can be used to mitigate the reservoir presence and deliverability risk in exploratory and appraisal projects that lack of reliable seismic inversion data.

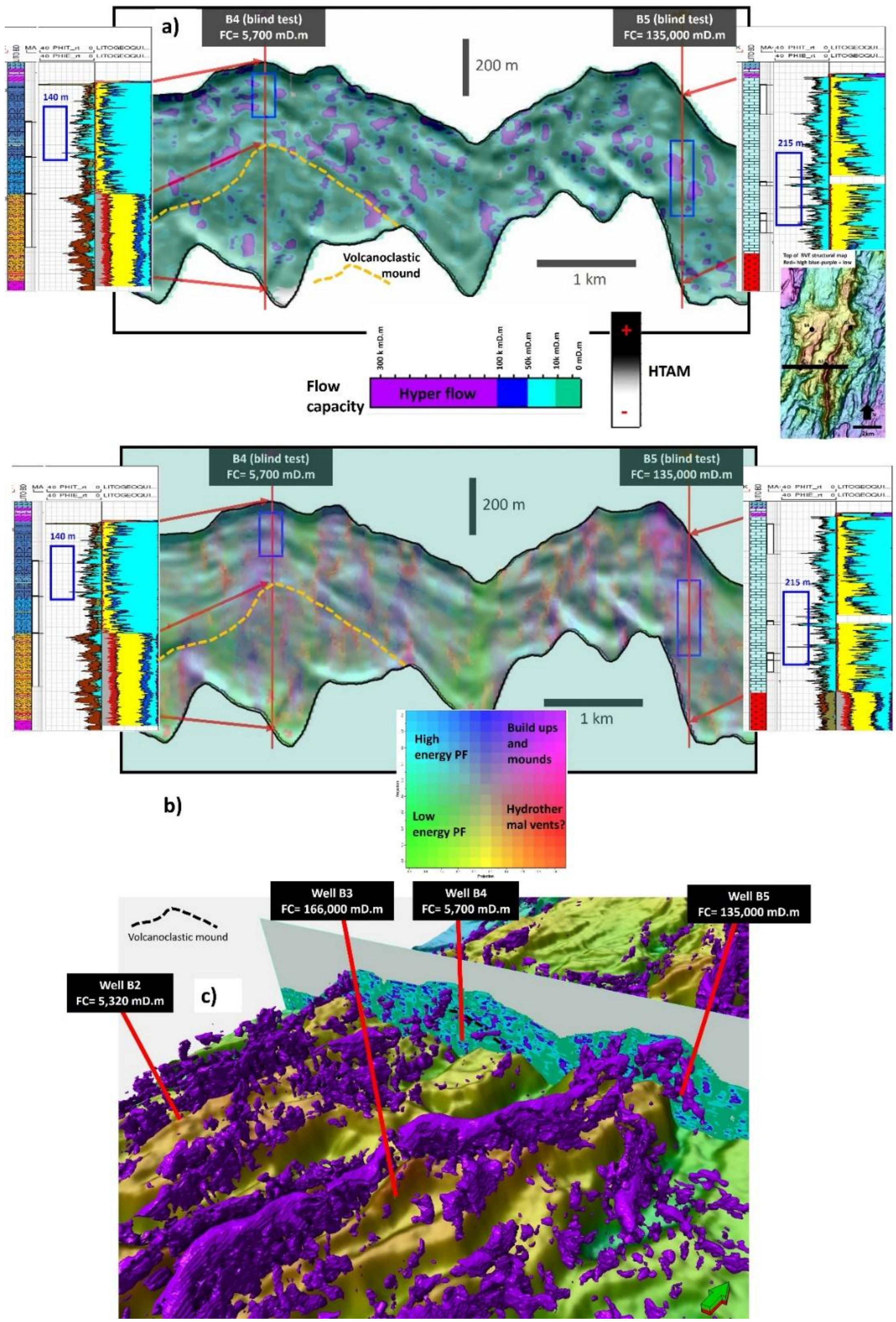

Figure 4.14- Transfer learning from Mero to Bacalhau field: a) Predictive model based on random forest regression (RFR) of flow capacity (FC) at BVE reservoir at the Bacalhau field using local HTAM, AVF, ERS, K1, K2, and GLET seismic

attributes inputs. We used the hyperparameters (400 regression trees, max depth=10) of the best RFR model that fitted the Mero field training dataset with hyper productive wells (Mero high) and built a predictive model of FC. At the seismic section passing through B4 and B5 wells from Bacalhau field, we co-rendered HTAM (seismic bandwidth pseudo impedance) and RFR flow capacity prediction. b) Same as A but the overlay is the SOM clustering using the same attributes. c) geobodies from SOM clusters related to the build ups and a vertical section of the RFR of flow capacity model.

Lapa field application

The predictive model results are shown in Figures 4.15 and 4.16. The random forest model of FC (Figures 15a and c) was able to correctly predict the values at perforated zones in 3 out of 4 blind test wells (75% of performance).

In the Lapa Field, we ran the RFR of FC using SWEET, HTAM, ERS, K1, and GLEN seismic attributes as input features. We used the same hyperparameters that fit the low productivity Mero training subset (Mero low) due to its more geological similarity to the Lapa field. Because of attribute importance change due to differences between the Mero high and Mero low subsets (mainly the absence of hyper flow values in the Mero low subset), we dropped K2 and GLET attributes and used SWEET and GLEN instead (Figure 4.12b). Despite their low importance for FC prediction in RFR, we kept at least one of the GLCM attributes because they are still responsible for most of the variability observed in PCA (Figure 4.6). In this case, we kept the GLCM energy one (GLEN) because of its higher importance for the RFR of FC compared to GLET and GLEH. A slight increment in RFR hyperparameters was needed to fit the model to the Mero low training data subset, as the best prediction in GSCV (R2= 83%) was achieved with 500 regression trees with a maximum depth of 12, a little higher than the Mero high training (400 regression trees, max depth = 10). In the Lapa field, the highest FC value is of the order of 20,000 mD.m, and it was measured in reworked facies represented by dolomitized grainstones and rudstones with shrubstones (stromatolites) fragments transported from nearby mounds (Well L2, Figures

4.7, 4.8a, 4.15 and 4.16). Because of the Well L2 fault proximity, the higher degree of dolomite cement dissolution was responsible for the flow capacity increment when compared to background values in Wells L1, L3, and L4 (between 5,000 and 6,000 mD.m) related to strong dolomitization. In Figure 4.15a, the higher FC values can be interpreted as either preserved depositional permeability or enhanced permeability due to dissolution closer to the faults. SOM clustering was able to depict the strongly dolomitized, higher impedance, and extremely low permeability intervals (strong positive HTAM values, reddish-pinkish clusters in Figure 4.15b). The higher permeability intervals are related to the hydrothermal zones close to the faults and depositional permeability were also depicted in different and specific SOM clusters. Figure 4.15c shows a 3D view of the highest FC and the SOM clusters geobodies related to the sparse mounds and a middle BVE surface with extracted ERS values. We can see that higher FC values are often spatially associated with the major faults, which can be explained by the higher intensity of dolomitic cement dissolution due to hydrothermal activity. In Figure 4.16, it is shown that both RFR and SOM clustering were able to correctly depict the different PLT flow units observed at DST perforations in wells L4 and L2. These two wells have the highest FC values in the Lapa field, and both are located close to major faults. In Figures 4.16a and 4.16b, the upper flow units are associated with more negative HTAM anomalies, which are compatible with the higher degree of hydrothermal dissolution observed in the core samples thin sections (Petrobras internal reports). The higher FC flow units are related to the greenish clusters, whereas the lower FC flow units are depicted as purple clusters related to strong dolomitization. In Figures 4.16c and 4.16d, the PLT flow units and SOM clusters association follow the same pattern: higher FC, lower HTAM and green clusters (higher dissolution), while lower FC flow units are associated with higher HTAM anomalies and purple clusters (strong dolomitization).

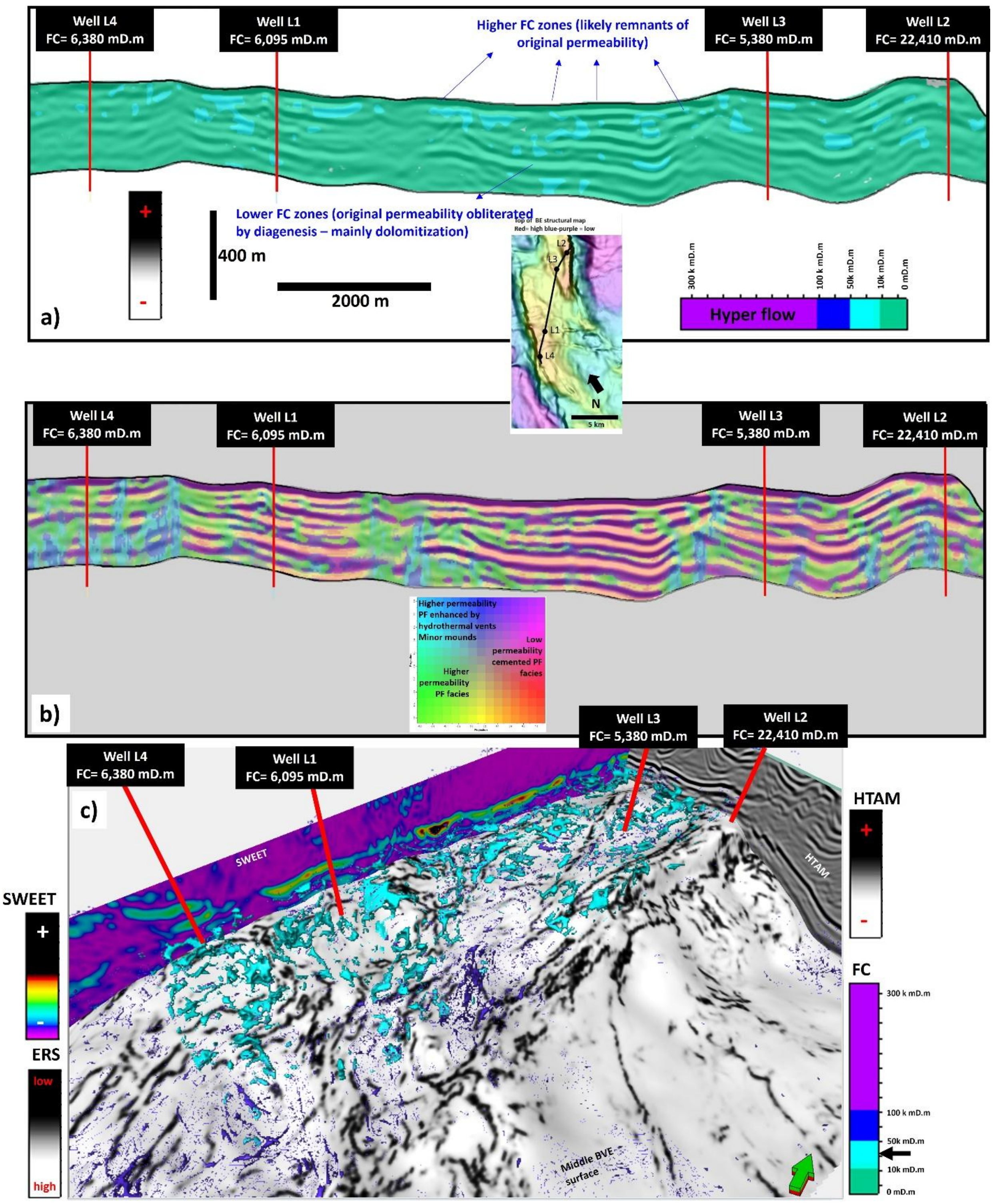

Figure 4.15- Transfer learning from Mero to Lapa field: a) Predictive model based on random forest regression (RFR) of flow capacity (FC) at BVE reservoir at the Lapa field using local HTAM, SWEET, ERS, K1, and GLEN seismic attributes inputs. We used the hyperparameters (500 regression trees, max depth=12) of the best RFR model that fit the Mero field training dataset without hyper productive wells (Mero low) and built a predictive model of FC. At the seismic section

passing through all studied wells from Lapa field (a), we co-rendered HTAM and RFR regression model. b) Same as 15a but the overlay is the SOM clustering model using the same seismic attributes. c) 3D view of the highest FC values (>10,000 mD.m) and SOM mounds geobodies over the middle BVE surface co-rendered with the ERS attribute. Some of the higher FC blobs are associated with faults (dissolution) whereas others are faraway (original permeability remnants).

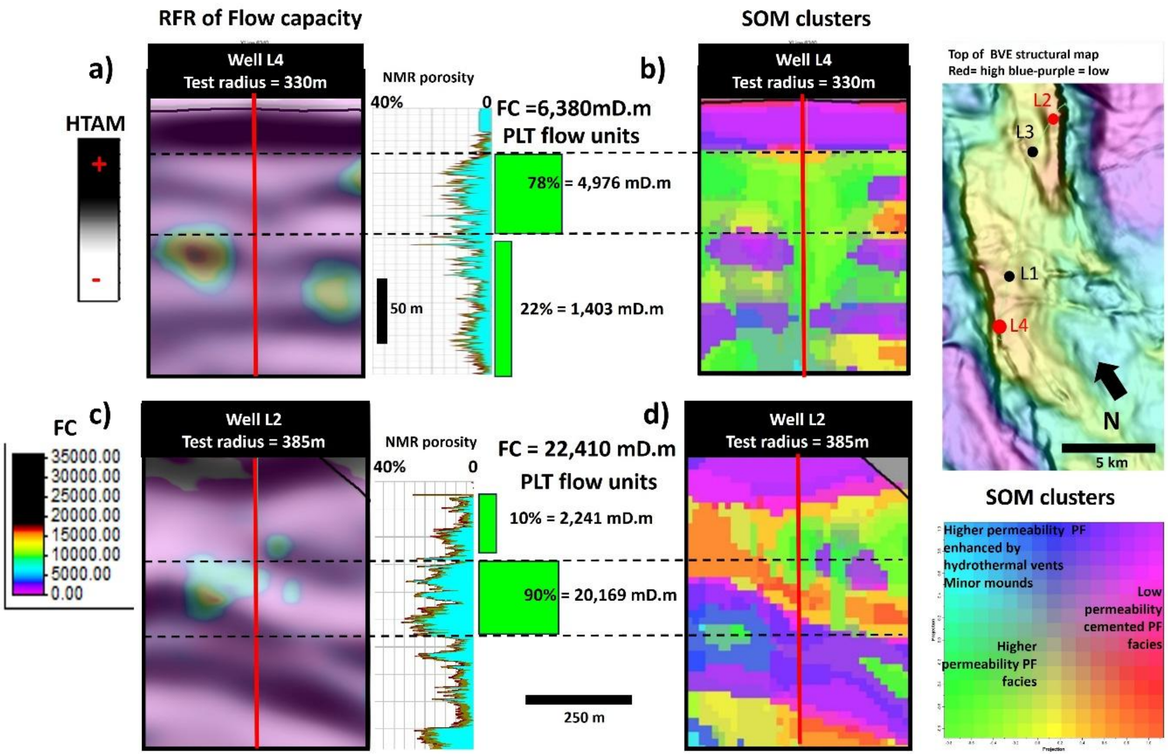

Figure 4.16- Detailed results of RFR of FC and SOM clustering around two of the wells studied in the Lapa field. a) co-rendered HTAM and predicted FC along the perforated interval at Well L4 with the NMR porosity log and PLT flow units. b) SOM clusters around the same well. c) co-rendered FC prediction and HTAM attributes around Well L2 perforated zones. d)SOM clusters around the same well.

Insights and implications

Our overall blind test performance of TL using Mero training in the target areas is 80%. It is important to mention that this is not a lithology classification task, it is a supervised regression of an important reservoir test parameter (flow capacity) that has a very high correlation with the reservoir permeability and commercial daily production performance. Moreover, our random forest algorithm was not trained with FC values coming from volcanoclastic mound perforations (which do not exist in the source area), it was trained instead with zero FC values from near-zero porosity sill intrusions close to perforated zones at BVE reservoirs from the Mero field (Figure 4.10). That is an indication that supervised regression transfer learning is more flexible than classification, as it would be very unlikely that a random forest classification model could correctly identify a volcanoclastic mound without being specifically trained for it. The SOM clustering at the Bacalhau field using the same attributes of Mero and the same training parameters, was able to depict the facies variation observed at the wells, including the volcanoclastic and carbonate mounds discrimination (Figure 4.14b). At the Lapa field, we got a 75 % prediction performance (3 out of 4 correct predictions at perforated zones, including PLT flow units, Figures 4.15a and 4.16). The SOM clustering using the same attribute combination could also correctly separate the higher and lower flow capacity flow units in different clusters. Considering we are working with post-stack seismic attributes rather than sophisticated seismic inversion attributes as predictors, and we are transferring the learning obtained from another dataset hundreds of kilometers far away, it is an outstanding performance. Figure 4.17 suggests that despite not being in the very top rank of RFR and SHAP attribute importances (Figure 4.12a and 4.12b), energy ratio similarity (ERS) is an especially important attribute for flow capacity prediction using random forest regression. In the three different areas, the highest FC values are spatially associated with low similarity zones related to either hydrothermal karstification inside mounds and build ups (Mero and Bacalhau fields) or dolomite cement dissolution along major faults in the Lapa field. The high performance of transfer learning from the Mero field to the other seismic surveys from the Bacalhau and Lapa fields indicates that this is a powerful reservoir de-risking tool that can be applied in other areas in the Santos Basin, including new exploration blocks with poor or even no well control. As we are working with transfer learning assisted flow capacity prediction using both production and injection tests for petroleum exploration and production optimization, this approach can be used for any kind of subsurface fluids management project like carbon capture, usage, and storage (CCUS), geothermal energy, and hydrogen geological storage.

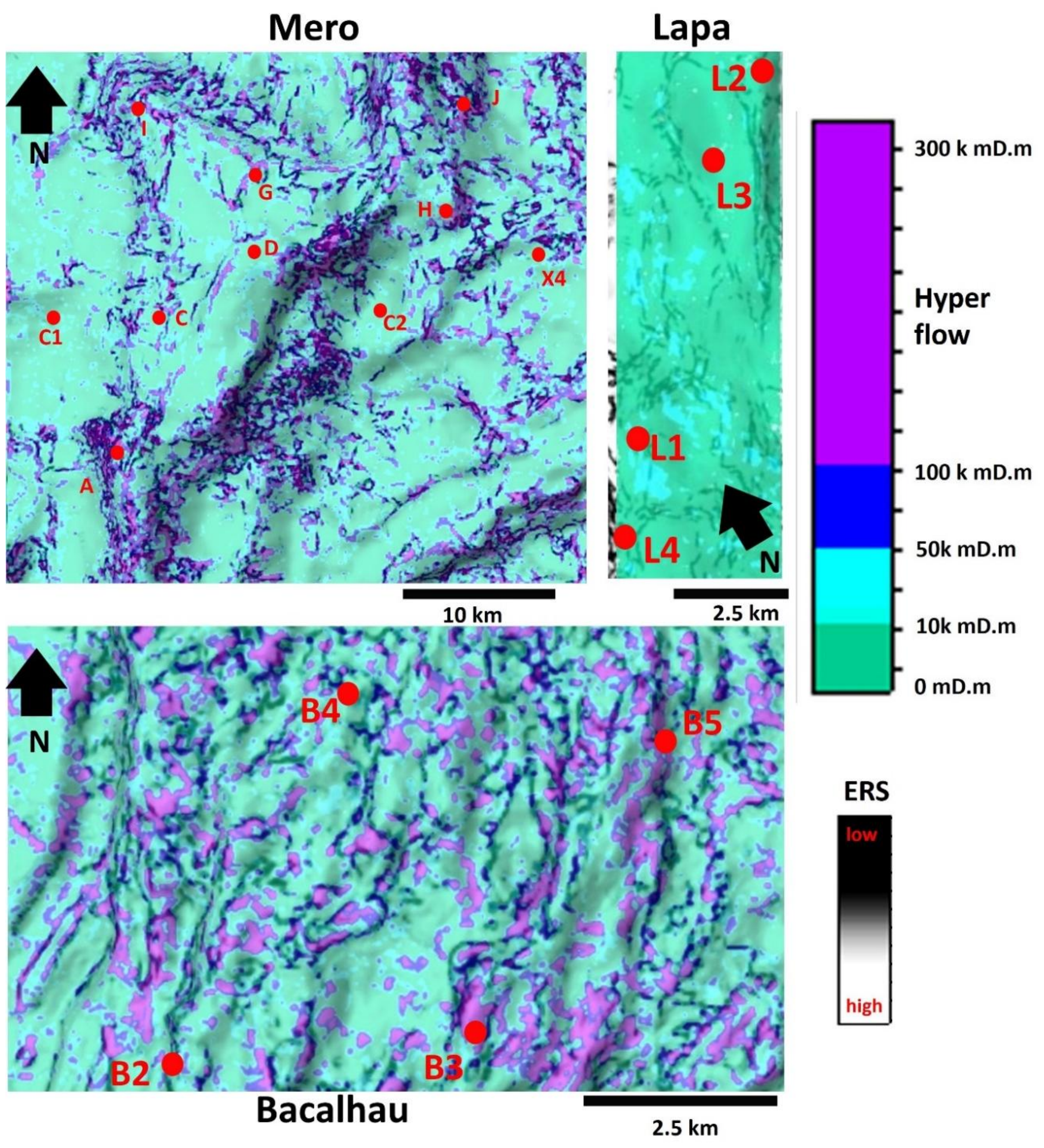

Figure 4.17- Map view of horizon probes of an intra BVE interval (middle of the reservoir) with random forest regression predicted FC values co-rendered with ERS attribute, one of the most important for FC prediction. At the Mero and Bacalhau fields, we can see that the hyper flow units (FC>100,000 mD.m) associated with strongly karstified mounds and build ups are closely associated with lower similarity anomalies (faults, fractures and karst), whereas the lower values are far away from these features. At the Lapa field, the higher FC values are also associated with the faults due to the higher degree of dissolution, but the overall flow capacity is very low.

Conclusions

Random forest regression (RFR) and SOM clustering algorithms trained and validated with Mero field (TL source area) seismic attributes and well dynamic data, accurately predicts flow capacity and reservoir facies at seismic scale in blind test locations at the Bacalhau and Lapa fields (TL target areas) with prediction performances of 85% and 75% respectively, even these fields are covered by different seismic surveys. Due to the high similarity between Mero and Bacalhau areas, no feature engineering was required. We simply transferred the hyperparameters learned at the Mero Field. In Lapa field, after exhaustive feature engineering (training dataset and attributes manipulation) and hyperparameter optimization, we got a 75% prediction performance. In both target areas, SOM clustering using the same input seismic attributes combination and training parameters from the Mero field was able to accurately depict facies distribution. For transfer learning across the three different seismic surveys, the most important seismic attributes were AVF, K1, K2, ERS, SWEET, HTAM, and GLET.

Because of the high prediction performance of transfer learning (80% on average), RFR of flow capacity can be used as a reservoir de-risking tool in earlier-stage exploration areas in the Santos Basin, where post-stack attributes are easy to obtain and there is poor or no well control. This approach can be applied to any kind of subsurface fluids management project (CCUS, geothermal, and hydrogen storage) in the context of energy transition for upstream optimization.

Acknowledgements

We would like to thank Petrobras and ANP (Agencia Nacional do Petroleo) for public data release for academic research. Thanks to Petrobras for financial support for the first author's PhD research at the University of Oklahoma School of Geosciences, which provided a very pleasant place for scientific research and academic learning. We are grateful to SLB for Petrel software licenses for OU Geosciences. Finally, many thanks to OU AASPI research group colleagues for sharing ideas and friendship.

References

Bäck, J. 2019. Domain similarity metrics for predicting transfer learning performance, Dissertation.

Bozinovski, S., Fulgosi, A. 1976. The influence of pattern similarity and transfer learning upon training of a base perceptron B2" (original in Croatian), Proceedings of Symposium Informatica 3-121-5.

Bozinovski, S., Santic, A., Fulgosi ,A. 1977. Normal teaching strategy in pair-association in the case teacher:human-learner:machine. (original in Croatian: Normalna strategija obicavanja u obucanju asocojacije parova u slucaju ucitelj:covjek-ucenik:masina), Proc. Conf. ETAN, 21:IV-341-346, Banja Luka, [available online].

Bozinovski, S. 1985. Adaptation and training: A viewpoint. Automatika 26 (3-4) 137-144

Bozinovski, S. 2020. Reminder of the First Paper on Transfer Learning in Neural Networks, 1976. Informatica. 44. 10.31449/inf.v44i3.2828.

Caruana, R. 1997. Multitask Learning, Ph.D. Thesis, School of Computer Science, Carnegie Mellon University.

Chai, C., Maceira, M., Santos- Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W.

2020. Using a deep neural network and transfer learning to bridge scales for seismic phase picking. Geophysical Research Letters, 47, e2020GL088651. https://doi.org/10.1029/2020GL088651

Chopra, S., Marfurt, K.J., 2007. Seismic Attributes for Prospect Identification and Reservoir Characterization: Society of exploration geophysicists.

Civitarese, D.S., Szwarcman, D., Gama e Silva, R.G., Vital Brazil, E. 2018. Transfer Learning Applied to Seismic Images Classification. doi: http://dx.doi.org/10.1306/42285Chevitarese2018.

Correa, R. S. M. 2023. Structural diagenesis and spatial arrangement of fractures in carbonate rocks. Doctoral dissertation. University of Texas at Austin.

Cunha, A., Pochet, A., Lopes, H., Gattass, M. 2020. Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data, Computers & Geosciences, Volume 135, 104344, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2019.104344.

Dong, Y., Zhang, Y., Liu, F., Cheng, X. 2021. Reservoir Production Prediction Model Based on a Stacked LSTM Network and Transfer Learning. ACS Omega. XXXX. 10.1021/acsomega.1c05132.

Ehrig, C., Sonnleitner, B., Neumann, U., Cleophas, C., & Forestier, G. (2024). The impact of data set similarity and diversity on transfer learning success in time series forecasting. *arXiv preprint arXiv:2404.06198*.

El Zini, J., Rizk, Y., Awad, M. 2019. A Deep Transfer Learning Framework for Seismic Data Analysis: A Case Study on Bright Spot Detection. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-11. 10.1109/TGRS.2019.2950888.

Gomes, Joao & Bunevich, Rodrigo & Tedeschi, L.R. & Tucker, Maurice & Whitaker, Fiona. (2019). Facies classification and patterns of lacustrine carbonate deposition of the Barra Velha Formation, Santos Basin, Brazilian Pre-salt. Marine and Petroleum Geology. 113. 104176. 10.1016/j.marpetgeo.2019.104176.

Ha, Thang & Lubo-Robles, David & Marfurt, Kurt & Wallet, Bradley. (2021). An in-depth analysis of logarithmic data transformation and per-class normalization in machine learning: Application to unsupervised classification of a turbidite system in the Canterbury Basin, New

Zealand and supervised classification of salt in the Eugene Island mini-basin, Gulf of Mexico. Interpretation. 9. 1-109. 10.1190/int-2021-0008.1.

Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441, 498–520. (doi:10.1037/h0071325)

Huang, W., Tian, Y. 2024. Reservoir parameters prediction based on spatially transferred long short-term memory network. PLoS ONE 19(1): e0296506. https://doi.org/10.1371/journal.pone.0296506.

Jolliffe, I.T., Cadima J. 2016. Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A 374: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202.

Kangxu, R., Oliveira, M.J., Zhao, J., Zhao, J., Oliveira, L., Rancan, C., Carmo, I., Deng, Q. 2019. Using Wireline Logging and Thin Sections to Identify Igneous Contact Metamorphism and Hydrothermal Influence on Presalt Limestone Reservoirs in Libra Block, Santos Basin. 10.4043/29818-MS.

Lima, B. E., Tedeschi, L., Pestilho, A. L., Santos, R., Vazquez, J., Guzzo, J., De Ros, L. 2020. Deep-burial hydrothermal alteration of the Pre-Salt carbonate reservoirs from northern

Campos Basin, offshore Brazil: Evidence from petrography, fluid inclusions, Sr, C and O isotopes. Marine and Petroleum Geology. 104143. 10.1016/j.marpetgeo.2019.104143.

Lubo-Robles, D., D. Devegowda, V. Jayaram, H. Bedle, K. J. Marfurt, and M. J. Pranter. 2022. Quantifying the sensitivity of seismic facies classification to seismic attribute selection: An explainable machine-learning study: Interpretation, 10, no. 3.

Lundberg, S., S. Lee. 2017. A unified approach to interpreting model predictions: Proceedings of the 31st Conference on Neural Information Processing Systems, 4765–4774.

Lundberg, S. M., G. G. Erion,S. Lee. 2018. Consistent individualized features attribution for tree ensembles: arXiv preprint arXiv:1706.06060.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. Lee. 2020. From local explanations to global understanding with explainable AI for trees: Nature Machine Intelligence, 2, 56–57, doi: 10.1038/s42256-019-0138-9.

Maas, M. V. R., Bedle, H., Matos, M. C. 2023.Seismic identification of carbonate reservoir sweet spots using unsupervised machine learning: A case study from Brazil deep water

Aptian pre-salt data. Marine and Petroleum Geology, Volume 151, https://doi.org/10.1016/j.marpetgeo.2023.106199.

Maas, M. V. R., Bedle, H., Ballinas, M.R., Matos, M. C. 2024. Unraveling the hidden relationships between seismic multiattributes, well-dynamic data, and Brazilian pre-salt carbonate reservoirs productivity: A shallow versus deep machine-learning approach. Interpretation 12(4):1-61. https://doi.org/10.1190/int-2023-0113.1

Mimoun, J.G., Fernández-Ibáñez, F. 2023. Carbonate excess permeability in pressure transient analysis: A catalog of diagnostic signatures from the Brazil Pre-Salt, Journal of Petroleum Science and Engineering, Volume 220, Part A, 111173, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2022.111173.

Molnar, C. 2021. Interpretable machine learning: A guide for making black box models explainable, https:// christophm.github.io/interpretable-ml-book/shap.html, accessed 11 August 2021.

Moreira, J.L.P., Madeira, C.V., Gil, J.A., Machado, M.A.P. 2007. Bacia de Santos. Rio de Janeiro. Boletim de Geociências da Petrobras, v. 15, n. 2, 531-549.

Oliveira L. C., Rancan, C. C. (2018): Ocorrência de condutos hidrotermais (hydrothermal vents) na seção aptiana da Bacia de Santos. In Congresso Brasileiro de Geologia, 49, Rio de Janeiro, 2018, p. 2008.

Oliveira, L.C., Rancan, C.C., 2019. Sill emplacement mechanisms and their relationship with the Pre- Salt stratigraphic framework of the Libra Area (Santos Basin, Brazil). LASI VI-The physical geology of subvolcanic systems: laccoliths, sills and dykes. Malargüe, Argentina.

Oliveira, Leonardo Costa de and Penna, Rodrigo Macedo and Rancan, Cristiano Camelo and Carmo, Isabela de O. and Marins, Gabriel Medeiros, Seismic Interpretation of the Mero Field Igneous Rocks and its Implications for Pre- and Post-Salt Co2 Generation – Santos Basin, Offshore Brazil. Available at SSRN: https://ssrn.com/abstract=4394061 or http://dx.doi.org/10.2139/ssrn.4394061

Pan, S. J., Yang,Q. 2010. A Survey on Transfer Learning, in *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, doi: 10.1109/TKDE.2009.191.

Pearson K. 1901. On lines and planes of closest fit to systems of points in space. Phil. Mag. 2,559–572. (doi:10.1080/14786440109462720).

Poros, Z., Jagniecki, E., Luczaj, J.,Kenter, J., Gál, B., Correa, T., Diniz Ferreira, E. L., Heumann, M., Elifritz, A., Johnston, M., Mcfadden, K., Matt, V. 2017. Origin of Silica in Pre-Salt Carbonates, Kwanza Basin, Angola. 10.13140/RG.2.2.31863.42405.

Rawat, T., Khemchandani, V. 2019. Feature Engineering (FE) Tools and Techniques for Better Classification Performance. International Journal of Innovations in Engineering and Technology. 10.21172/ijiet.82.024.

Sartorato, A.C.L., Tonietto, S.N., Pereira, E., 2020. Silicification and Dissolution Features in the Brazilian Presalt Barra Velha Formation: Impacts in the Reservoir Quality and Insights for 3D Geological Modeling. Rio Oil & Gas, Rio de Janeiro, Brazil. https://doi.org/10.48072/2525-7579.rog.2020.068.

Stica, J. M.; Zalan, P. V.; Ferrari, A. L., 2014. The evolution of rifting on the volcanic margin of the Pelotas Basin and the contextualization of the Paraná-Etendeka LIP in the separation of Gondwana in the South Atlantic. *Marine and Petroleum Geology*, [Amsterdam], v. 50, p. 21.

Thrun, S., Pratt, L.Y. 1998. Learning to Learn: Introduction and Overview. Learning to Learn.

Tritlla, J. & Loma, R., Sanchez Perez-Cejuela, V., Cerdà, M., Wloszczowski, D., Perona, R., Algíbez, J., Caja, M. A., Levresse, G., Bonillo, P., Tiwary, D. ,2013. Volcanism, alkalinity and hydrothermalism in offshore Brazil: unraveling the Pão de Açúcar reservoir. Repsol E&P Conference Madrid, Spain.

Wong, L. J., McPherson, S., and Michaels, A. J. 2022. Quantifying Raw RF Dataset Similarity for Transfer Learning Applications, in *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2076-2086, doi: 10.1109/OJCOMS.2022.3218502.

Yan, K., Zuo, Y., Zhang, Y., Yang, L., Pang, X., Wang, S., Li, W., Song, X., Yao, Y. 2024. A study on the accumulation model of the Santos basin in Brazil utilizing the source–reservoir dynamic evaluation method. Scientific Reports. 14. 10.1038/s41598-024-69756-y.

Yang, J., Lin, N., Zhang, K., Ding, R., Jin, Z., Wang, D. 2023. A data-driven workflow based on multi-source transfer machine learning to gas-bearing probability distribution prediction: A case study. GEOPHYSICS. 88. 1-81. 10.1190/geo2022-0726.1.

Yang, J., Lin, N., Zhang, K., Fu, C., Zhang, C. 2024. Transfer learning-based hybrid deep learning method for gas-bearing distribution prediction with insufficient training samples and uncertainty analysis, Energy, Volume 299, 131414, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2024.131414.

Zalan, P. V. 2014. Comparisons between the Deep Crustal Structure of Magma-Poor and Volcanic Passive Margins. In: Atlantic Conjugate Margins Conference, 4. St. John's, Canada. *Extended Abstracts…* St. John's, Canada: [S.n.], 2014. p.75–78.

Zalan, P. V. 2024. Remaining Potential Brazil DERBYANA 2024. Derbyana. 45. 60 pages. 10.14295/derb.v45.826.

Chapter 5: Conclusions

In this dissertation, I systematically studied how post stack seismic multi-attributes and dynamic data from production assets can be analyzed together via machine learning to make accurate predictions of subsurface reservoir productivity and injectivity for upstream projects optimization. During the upstream phase, the only seismic data available are post-stack attributes. Dozens of wells are required to obtain reliable seismic impedance volumes, which do not exist in exploration assets. As mentioned before, the main research question that motivates my PhD research is: **"Do post stack seismic attributes significantly predict reservoir facies distribution and well productivity during upstream phase?"**

The answer is definitely yes. In my first research project (Chapter 2), I started with unsupervised learning assisted reservoir modelling using multiple seismic attributes (HTAM, AVF, ERS, K1, K2, GLET and VEL) and clustering algorithms (SOM, GTM, and K-means). I found that all three methods were able to automatically build facies models that separated the most productive reservoir facies (carbonate mounds and build ups, the sweet spots) from less productive, dolomitized ones from lower energy platform facies, magmatic sills (flow and drill barriers), and distal mudstones (non-reservoirs). K-means clustering discriminates the sweet spots and igneous intrusions but could not separate productive from non-productive reservoirs of the platform and distal facies respectively. However, K-means clustering took one fifth of the computational time (five times faster) when compared to SOM and GTM and can be a good option if it is the only one available. All models were validated using well information from the Barra Velha (BVE) carbonate reservoir from Mero field, which has a plethora of reservoir and production data. My models could also identify several appraisal candidates in an earlier exploration stage area (Libra) adjacent to the Mero field and covered by the same seismic survey. My first research project proved that unsupervised clustering and seismic multi-attributes are excellent tools for qualitative reservoir modelling in areas close to production assets to optimize the appraisal phase once they map the most critical lithologies (reservoir sweet spots, flow and drill barriers) quickly. Unsupervised clustering is an excellent approach for automatic 3D seismic attributes assisted reservoir mapping that provides, in hours, models compatible with the human-constructed ones that can take months to build. The main limitation is that as the seismic attribute choice is a human task, it can be a source of bias. Additionally, it is a qualitative approach.

In the second project (Chapter 3), I extracted the same seismic attributes but changing VEL for SWEET, from nine traces around DST perforations in the BVE reservoir in ten wells from the Mero field for the training dataset building. In this project, I discovered the relationships among seismic attributes, the flow capacity (FC), productivity index (PI) and daily oil production rates observed in the wells. For this, I used shallow and deep supervised learning algorithms to fit models of FC and PI to Mero field main reservoir well and seismic data. Using random forest (RF), K-nearest neighbors, support vector machine, and multi-layer perceptron (MLP-deep learning) regression of FC and PI, I built predictive models of BVE reservoir productivity that fits up to 85% (random forest model) of the blind test wells used for model validation. Random forest regression was the fastest and most effective method with blind test performances ranging from 80 to 85%. Deep neural networks (MLP) provided less accurate models (75 % of blind test performance) at a much higher computational cost (30 times slower than random forest). The best predictive model was built with random forest regression of flow capacity using AVF, K1, K2, ERS, HTAM, and GLET as inputs, in this order of importance provided by both SHAP explanation and random forest importance. This model correctly predicted the BVE reservoir productivity in 17 out of 20 blind test wells and identified several appraisal candidates in the adjacent Libra area which is still in the appraisal phase and is covered by the same seismic survey. This shows that this supervised learning assisted quantitative interpretation is a powerful tool for reservoir derisking in near field exploration and appraisal projects. Like unsupervised clustering, it provides an automatic reservoir property mapping saving a lot of the human interpreter's time, which is one of the main advantages of artificial intelligence and machine learning. In terms of main limitations, this approach is valid only when the well test flow unit's thicknesses and seismic resolution match. Moreover, it was developed for exploration areas close to well-known production assets, covered by the same seismic survey.

In the third and last research project (Chapter 4), I explored if it is possible to use transfer learning by using the Mero training dataset machine learning to make reservoir productivity predictions in two other seismic surveys hundreds of kilometers far away from the Mero field but still in the same basin and same reservoir interval (BVE carbonates). I used the training hyperparameters of SOM and RFR reservoir models, the same seismic attributes, and transferred the learning to Bacalhau and Lapa field where there are some geological discrepancies that were handled by feature engineering in the training dataset. It was possible to make accurate predictions in the Bacalhau field, where there are non-productive volcanoclastic mounds discriminated by SOM and RFR as low productivity reservoirs (blind test performance of 85%). In the Lapa field, it was possible to correctly predict the lower productivity in 3 out of 4 blind test wells (75% blind test performance). For transfer learning across the three different seismic surveys, the most important seismic attributes were AVF, K1, K2, ERS, SWEET, HTAM, and GLET. Transfer learning proved a powerful derisking tool even when dealing with multiple seismic surveys, which can accelerate the upstream phase in the exploratory areas with poor or no well control. The main limitation is that the training data set is restricted to the Mero field, so, data augmentation is required to make it more diverse by including data points from other areas.

As a key scientific contribution, I developed an innovative and groundbreaking technique that utilizes big data from known assets and post-stack seismic attributes to create predictive models applicable across various upstream assets throughout the basin for project optimization.

Once I used both productivity and injectivity test data from the Mero field to train and validate my ML models, the presented methodology can be used in any kind of subsurface project like CCUS, geothermal and hydrogen storage using easy-to-obtain seismic attributes and optimize the upstream phase. I hope that this PhD research will be a reference for future work in other kinds of geological environments as onshore both conventional and unconventional reservoir ML assisted characterization for multiple purposes. I believe that the petroleum industry is a major global player with its subsurface operations expertise, in other words, it has the power to lead the energy transition by applying its vast data driven knowledge to low-carbon subsurface solutions and be a game changer in global warming mitigation. Count on me to collaborate with this.

Appendix 1 – State of art of petrography, petrogenesis and hydrothermal alteration of the Barra Velha Late Aptian lacustrine carbonates- Santos Basin, SE Brazil offshore.

Abstract

The highly productive Barra Velha Fm Late Aptian lacustrine carbonates of southeastern Brazilian basins hold dozens of billions of barrels of high-quality oil and are very different from those commonly observed at the literature. The main subject of disagreement among the geoscientists are the so called "microbial stromatolites". In fact, those stromatolite-like rocks are not made by any kind of biochemical components, they are mostly composed of fascicular calcite aggregates. The first researchers called them stromatolites due to their geometrical semblance to the biochemically built boundstones (algae, sponges, and coral stromatolites). The first petrographic classification schemes were based on Dunham (1962) and encountered some resistance by the scientific community due to the rareness of micrite, whereas the fine-grained material is mainly Mg-silicates. Another issue of debate is the either abiotic or abiotic origin of these lacustrine carbonates. Besides crenulated laminites, no other indication of microbial induced precipitation was found. Two other important characteristics of these reservoirs are the severe recrystallization and omnipresence of CO_2 in oil accumulations (ranging from traces up to 99% CO_2). After 2014, when the Brazilian Government released data for academic research, there was a big jump

on the understanding of petrogenesis, petrography and hydrothermal alteration of these lacustrine carbonates. A strong abiotic component driving the carbonate precipitation due to mantle degassing and crustal fluids circulation through the basin via deep faults was described using systematic petrography, fluid inclusions, and isotope data. Another abiotic process identified was the climate variations (humid x arid) and hydro geochemically controlled carbonate precipitation due to lake water evaporation. Petroleum companies' geoscientists also discovered a regional and pervasive hydrothermal silicification affecting a pre-existing karstic system with a net porosity gain at both Santos and Kwanza basins (conjugated rifted margins). Finally, recently published unifying petrographic classification schemes and a mixed origin (dominantly abiotic with minor biotic) for the carbonates is widely accepted and incorporated by the geoscientific community today. At the Santos-Kwanza conjugate margins, the main controls for the lacustrine carbonate precipitation are the lithospheric hyperextension with mantle exhumation and continuous degassing providing anions and cations via hot springs, which also lead to a regional hydrothermal alteration and locally high CO_2 content in the oil fields.

Introduction

Nineteen years after the discovery of the world class Brazilian pre-salt oil province in 2006, there is still an aura of mystery about the origin of the Late Aptian carbonate reservoirs of the Barra Velha formation. The petrographic classification itself of those "odd carbonates" is controversial. The first authors started with the classical Dunham scheme, but it failed due to the fact that the micrite matrix is often rare or absent (Gomes et al, 2020). With the evolution of exploration and new well data, the biogenic or abiotic origin of such rocks became the main theme of debate among the specialists. Another complicator is the pervasive hydrothermal alteration (mainly silicification) that affects the reservoirs in a positive way with pore system amplification (Tritla et al., 2013; Poros et al., 2017; Vieira de Luca et al., 2017; Sartorato et al., 2018, 2020; and Mimoun and Fernandez, 2023). During the second half of last decade, thanks to data releasing for academical purposes, a lot of studies about petrography, petrogenesis, and hydrothermalism were carried out and, thus, took the geological knowledge about the Barra Velha carbonates to another level. One important issue that is partially solved is the proposal of an adapted petrographic classification scheme due to the uniqueness of these reservoirs with an aim to unify the terminology. Another important aspect is the persisting dichotomy of the abiotic-biotic origin of the lacustrine carbonates. The main objective of this work is to show a compilation of the most prominent works carried out since 2006 until present day with an order to provide a state of the art and future trends of the geological knowledge about these “analog-less” high yield reservoirs. It is important to mention that the information presented here was not necessarily available at the time of its publication. Due to strong data confidentiality, most information took time to get from the industry to the academic community.

Geological Background

The Santos Basin is a passive margin basin which evolved from a continental rift started in the Early Cretaceous (Neocomian/Barremian). In terms of tectonostratigraphic evolution, the basin started after a mantle plume driven lithospheric stretching with the rift phase half grabens concentrated in the stretched and thinned continental crust domains during the Barremian until the Early Aptian (Zalan, 2014, Stica, 2014, and Zalan,2024). Within these grabens there are, from bottom to top: basic volcanic rocks, talc-stevensite sequences, lacustrine source rocks, and coquina deposits. In the Late Aptian, after the rift phase dominated by mechanical subsidence, a sag basin was developed with the deposition of lacustrine carbonate reservoirs of the Barra Velha (BVE) formation mainly in flat platforms over basement horsts and build ups and mounds related to deep faults (Figures AP1.1 and AP1.2). At the very end of the Late Aptian, when the oceanic stage has begun, a one to two kilometers thick evaporite sequence was deposited, followed by a typical transgressive marine drift sequence during the Late Cretaceous and Tertiary. During the Santonian, an alkaline magmatic event is responsible for the intrusion of mainly sills within the BVE reservoirs and overlying evaporites (Moreira et al, 2007 and Carvalho and Fernandes, 2021).

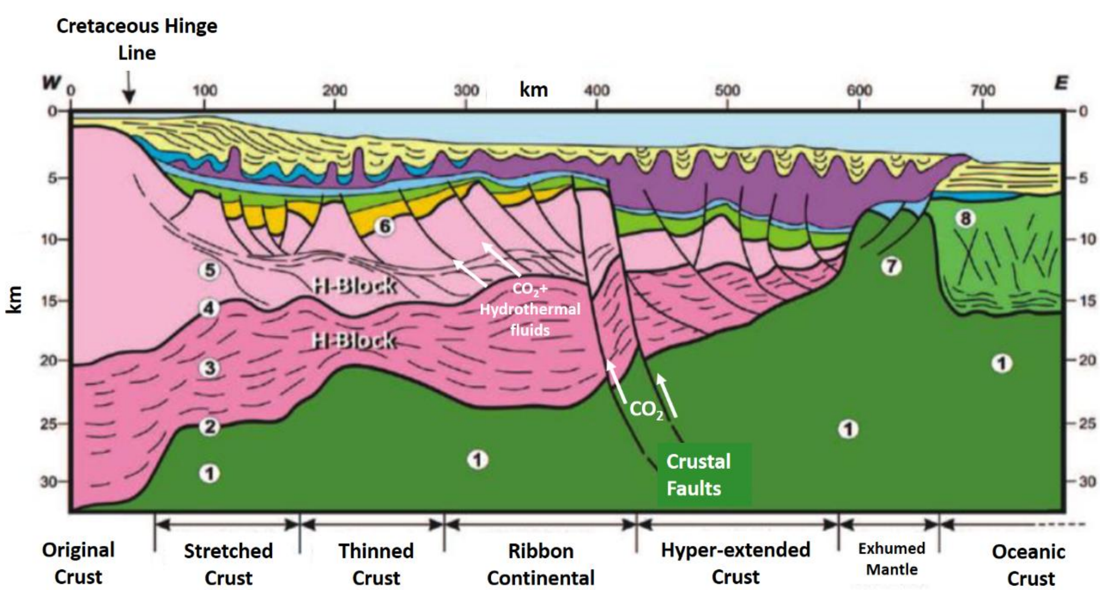

Figure AP1**.1 -** Crustal domains of the Santos Basin proposed by Zalan (2014) based on Stica (2014). The numbers indicate the: Mantle (1), Moho (2), Ductile Lower Crust (3), Conrad discontinuity (4), Brittle Upper Crust (5), Crystalline Basement top (6), Exhumed Mantle (7), and Oceanic Crust (8). Color code for the sedimentary units above the Crystalline Basement top are: Barremian Syn-Rift (yellow), Early Aptian Syn- Rift (green), Late Aptian Pos-Rift (light blue), Late Aptian evaporite deposits (purple), Post Ryft Albian Carbonates (dark blue), siliciclastic Late Cretaceous and Tertiary drift sequences (light yellow). The blue rectangle indicates the crustal segment of the Mero Field area, (modified from Zalan, 2014).

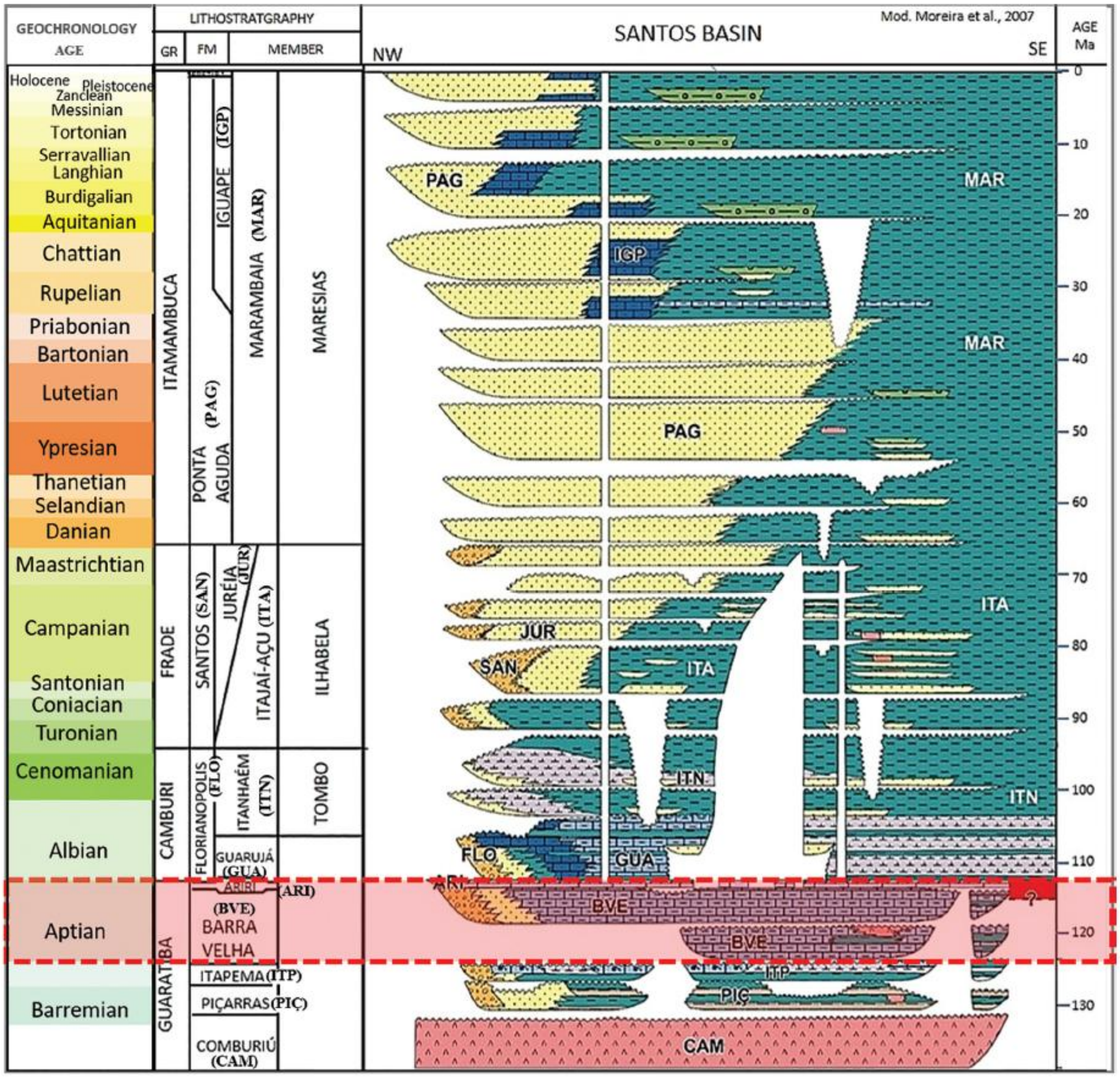

Figure AP1.2 **-** Santos Basin chronostratigraphic chart. The Barra Velha Fm reservoirs are highlighted by the red rectangle (Modified from Carvalho and Fernandes, 2021 and Moreira et al., 2007).

First proposed petrographic classification schemes and reservoir models

From first pre-salt carbonate well penetrations to first commercial production and world class play emergency (2002-2014)

In 2002, at the northern sector of Campos Basin, at Parque das Baleias (Whales Park) oil field cluster, where the main producing reservoirs are Late Cretaceous turbidites, Petrobras decided to drill deeper wells to investigate possible pre-salt Aptian reservoirs. Some wells penetrated very low permo-porosity oil bearing carbonates that were so called "microbialites". In 2006, 550 km to the southwest at the Santos Basin, the Tupi oil field wildcat well penetrated the same reservoirs, which have a perfect correlation at a basin scale (Carminatti et al 2018). (Figures AP1.3 and AP1.4). Dias et al., (2005), at the light of the concepts of Grotzinger and Knoll (1999), described them as a succession of microbial laminites, dendritic bush shaped stromatolites and laminated mudstones (Figures AP1.5A, AP1.5B and AP1.5C).

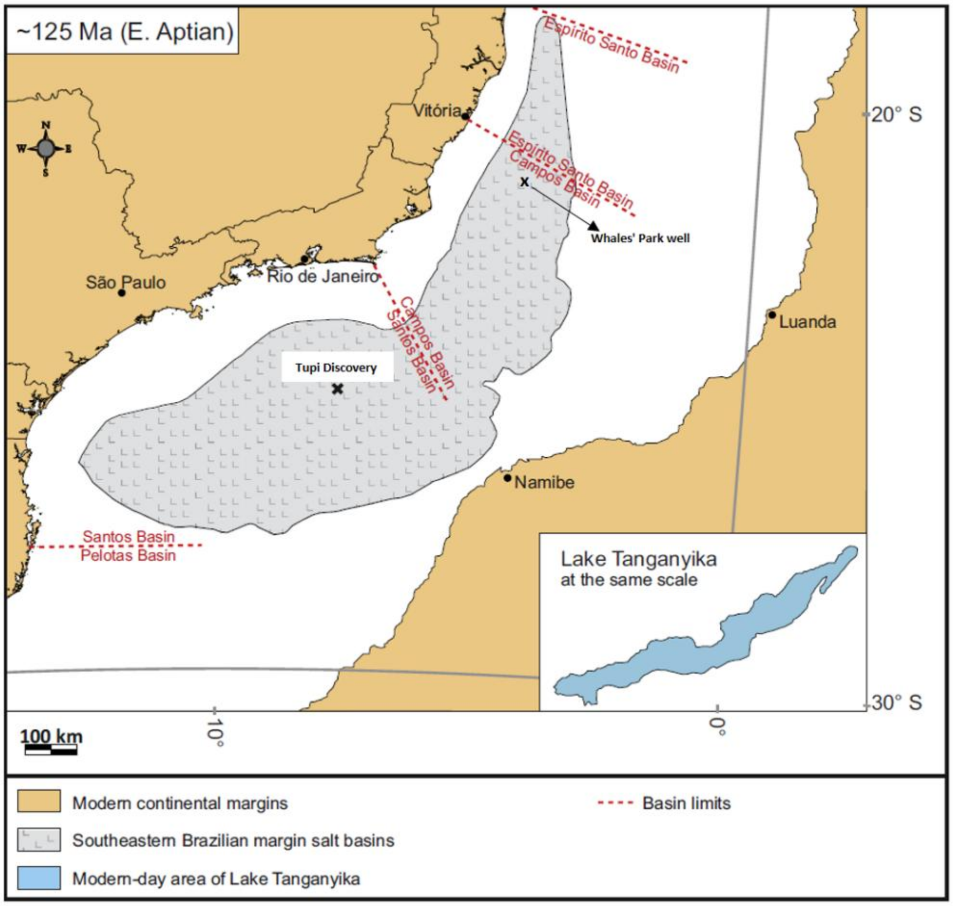

Figure AP1.3 - Early Aptian plate reconstruction of South Atlantic margins showing the extent of the pre-salt Aptian Lacustrine Carbonates and the position of the first two wells which drilled these reservoirs. (Modified from Pietzsch et al., 2018).

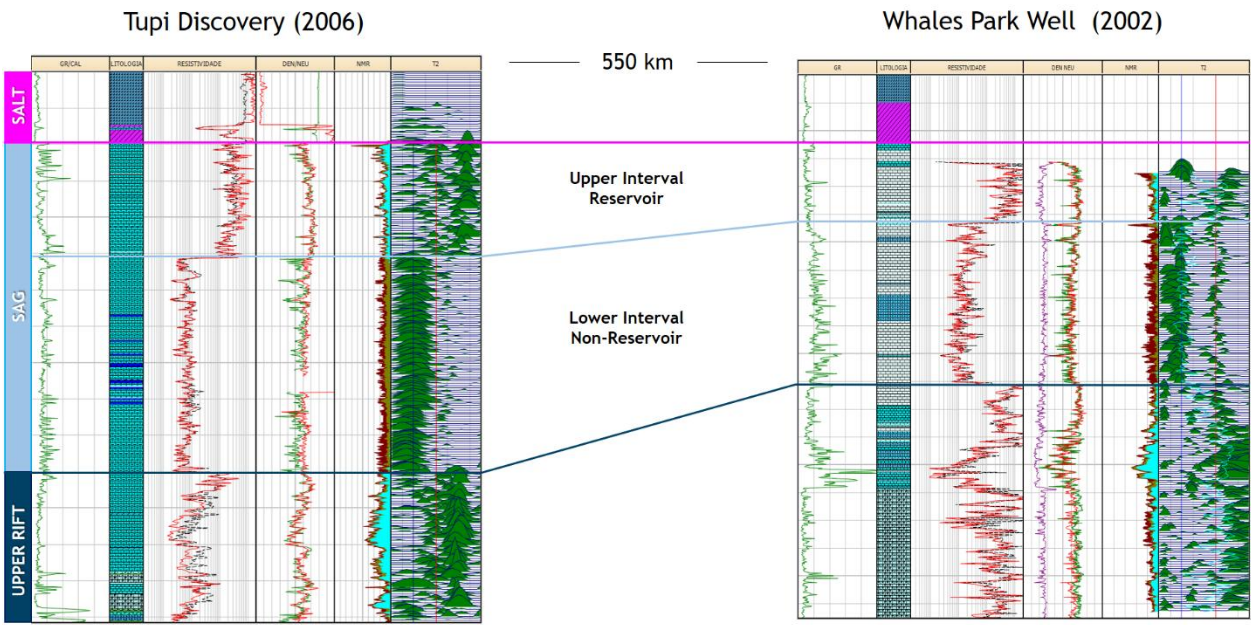

Figure AP1.4 - Regional scale well correlation of the Aptian pre-salt carbonates of Santos Basin (Tupi well) and Campos Basin (Whales Park well), see Figure. 3. (Modified from Carminatti et al., 2018).

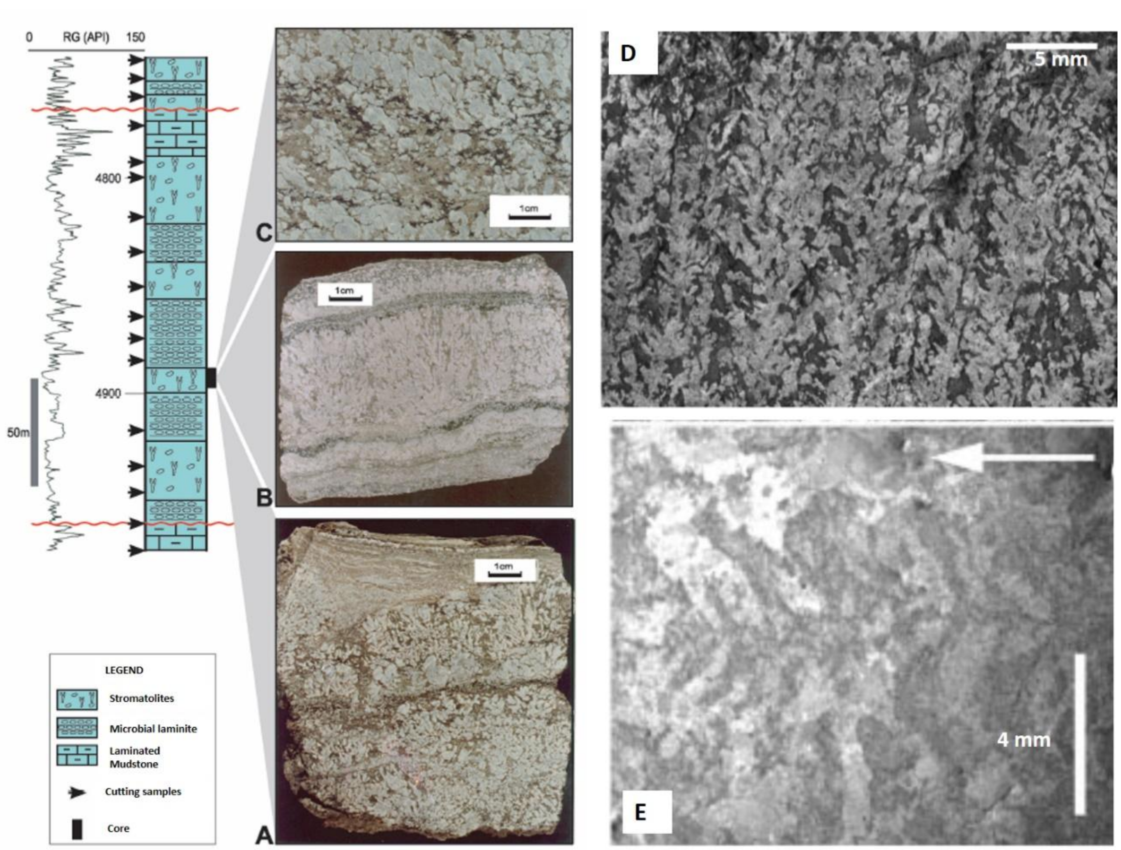

Figure AP1.5- Third order sequence proposal for Campos Basin Late Aptian carbonates (Dias et al., 2005). A) Dendritic bush shaped stromatolites; B) Dendritic stromatolites; C) Bush shaped stromatolites. This classification used the terminology proposed by Grotzinger and Knoll (1999): D) Outcrop photograph of dendritically branching tufa (light) which forms stratiform sheets on the scale of1–2 cm. Branching structures are infilled with marine cement and micritic sediment. E) Photomicrograph of a single dendrite structure with thin central stalk and numerous broad branches, both formed of micrite (dark). Lighter material is void-filling marine cement. Note arrow indicating growth orientation. (Modified from Dias et al., 2005 and Grotzinger and Knoll, 1999).

After data compilation from several wells drilled targeting the emerging pre-salt play at both Santos, Campos and Espirito Santo Basins (with multiple world class discoveries), Terra et al., (2010) proposed a classification scheme based principally on Dunham (1962) (Table AP1.1). Subsequent workers adopted this scheme (Tonietto et al., 2012) (Figure AP1.6). According to them, the facies distribution of microbiolites from BVE in the Basin of Santos was conditioned by paleo relief, where in the structural heights there were ideal conditions of water depth and little contribution of siliciclastic material, then, the development of these microbial bioherms was possible. But one problem was not totally addressed: the rareness or total absence of micrite in the system. Madrucci et al., (2013) carried out a detailed study of the fine-grained material of the Aptian carbonates and find out that they are mostly Mg-clay minerals (kerolite, stevensite, saponite and sepiolite). They proposed that the Mg-clay minerals are neo formed, authigenic, precipitated directly from ionic or colloidal solutions at a low energy, high salinity and alkaline subaqueous environment. The palaeontologic studies from Arai et al., (2014) confirmed a continental lacustrine environment for the Aptian carbonates of BVE at Santos Basin and their chrono correlates at the adjacent basins (Figure. AP1.7).

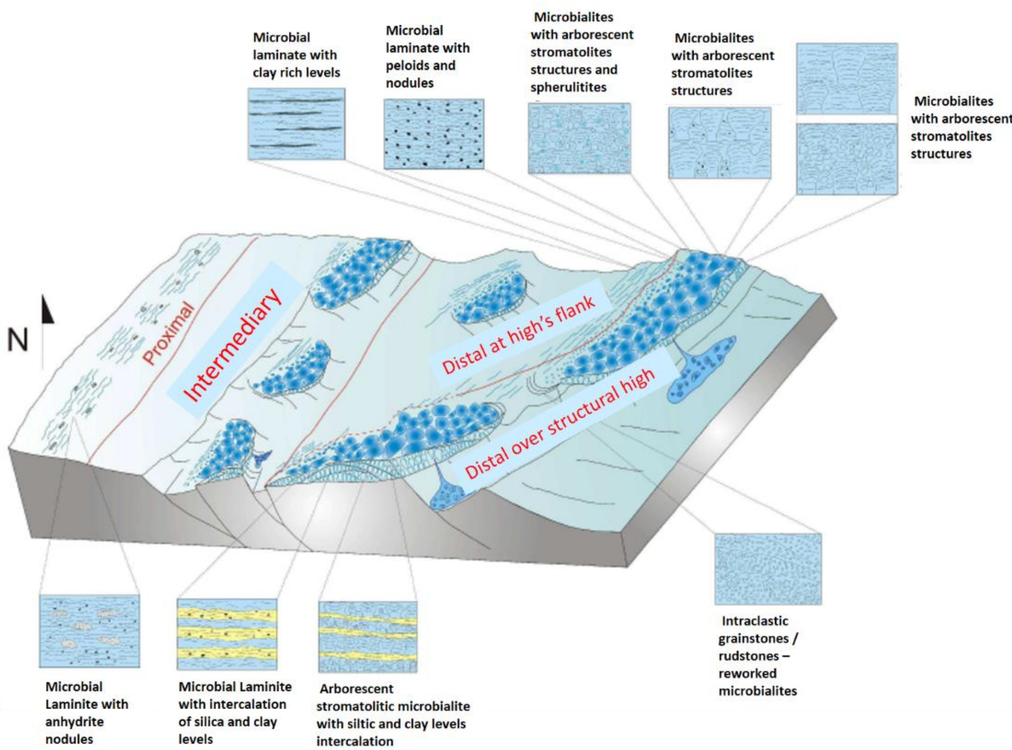

Figure AP1.6 - Depositional model for Barra Velha Fm carbonates at Santos Basin. The deposition of the "microbialites" was strongly controlled by the substrate paleo relief.

Carbonate Rock Classification - Applied to Brazilian Sedimentary Basins

	PRINCIPAL NOMENCLATURE	TEXTURE	DEFINITION	SUPPLEMENTARY NOMENCLATURE	REFERENCES	CORRELATE TERMS	MODIFIER TEXTURES	COMPONENT AVERAGE SIZES - VERTICAL SLICE AND PACKING	EXTERNAL SHAPE (core samples)
Components not bound during formation	MUDSTONE		Matrix supported - less than 10% grains (sand size or greater)	**WITH** (in case of presence of grains) oolites, oncolites, peloids, fecal pellets, intraclasts. bioclasts, spherulites or **BIRDSEYES CALCILUTITE**: matrix supported with fenestral porosity or spatic calcite lenses.	Dunham (1962)	Calcilutite, Dismicrite, Birdseyes Mudstone.			
	WACKESTONE		Matrix supported - more than 10% grains (sand size or greater)	**WITH** (for grains) oolites, oncolites, peloids, fecal pellets, intraclasts, bioclasts, spherulites.	Dunham (1962)				
	PACKSTONE		Grain supported - with matrix	"+" (principal composition): oolitic, oncolitic, peloidal, pelletoidal, intraclastic, bioclastic, spherulitic; with fragments of stromatolites, thrombolites, laminites, leiolites, dendrolite, spherulitites.	Dunham (1962)	Calcarenite, oomicrite, biomicrite, pelmicrite, intramicrite.			
	GRAINSTONE		Grain supported - no matrix (<5% matrix)	"+" (principal composition): oolitic, oncolitic, peloidal, pelletoidal, intraclastic, bioclastic, spherulitic, with fragments of stromatolites, thrombolites, laminites, leiolites, dendrolites, spherulitites.	Dunham (1962)	Calcarenite, oospatite, biospatite, pelespatite, intraspatite.	With exposure features; With pedogenetic features; Bioturbated; Dolomitized; Silicified, With Hardgrounds; etc.		
	FLOATSTONE		Matrix supported, more than 10% grains > 2mm.	**WITH** (for grains) - oolites, oncolites, peloids, fecal pellets, intraclasts, bioclasts, spherulites. With fragments of stromatolites, thrombolites, laminites, leiolites, dendrolites, spherulitites.	Embry&Klowan (1971)				
	RUDSTONE		Grain supported, more than 10% grains > 2mm.	"+" (principal composition): oolitic, oncolitic, peloidal, pelletoidal, intraclastic, bioclastic, spherulitic, with fragments of stromatolites, thrombolites, laminites, leiolites, dendrolites, spherulitites.	Embry e Klowan (1971)	Breccia			
	BIOACUMULATED		One type of organism predominates. No reworking (in situ) or greater.	"+" main organic component: such as ostracodes, bivalves, macroforams, crinoids, etc.	Carozzi (1972)	Coquina			
	BRECCIA		Grain supported rock. Grains: more than 50% angular and >2mm.	"+" (principal composition): intraclastic; with fragments of stromatolites, thrombolites, laminites, leiolites, dendrolites, spherulitites.	Flügel (2004); Pettijohn (1974)	Collapse breccia, Breccia, Rudstone with angular clasts.			
Original components bound together during formation - in situ	BOUNDSTONE		Original components bound together during deposition. Rock formed *in situ*	"+" main organic components: such as coralgal, rudists, stromatoporoids, etc...	Dunham (1962)	Biolithite, Bafflestone, Bindstone, Framestone.	With exposure features; With pedogenetic features; Bioturbated;Dolomitized; Silicified, etc		
	STROMATOLITE Laminated structure, generally convex; may show growth features/upward internal ramifications.Mostly microbial origins. Observed in different scales		**STROMATOLITE**						
			ARBORESCENT STROMATOLITE - internal components in divergent branching fashion. Lenght is greater than width.					Very Small: < 0,2 cm Small: 0,2-0,5 cm Medium: 0,51-1,5 cm Big> 1,5 cm **Dense:** all components touch each other; **Normal**: some components touch each other; **Open**: seldom or rare components touch each other	Finely laminated Slightly domic Domic Tabular Columnar Coneform
			ARBUSTIFORM STROMATOLITE - internal components may or may not branch out from the base. Ratio height/width aproximately 1:1						
			DENDRIFORM STROMATOLITE - internal components intensively branch in divergent way. Lenght is much longer than width.		Riding (2000)		Mottled Bioturbated With exposure features; With pedogenetic features; Dolomitized; Silicified, etc		
	THROMBOLITE		Macroscopically clotted texture, massive, domic. Mostly microbial origin.						Slightly laminated, Domic, Columnar
	DENDROLITE		Dendritic structural microbial deposit formed by skeletal cyanobacterias.						Slightly domic, Domic
	LEIOLITE		Domic microbial carbonates. Lamination and thrombs absent.						Slightly domic, Domic
	SPHERULITITE		Rock composed of spherical/subspherical particles, smooth or lobated rims (spherulites), generally smaller than 2mm, amalgamated or isolated.	Spherulite supported with clay (>10%) = SPHERULITITE WITH CLAY; Clay supported, with spherulites = ARGILLITE WITH SPHERULITES; Clay as lamellas= LAMELLAR ARGILLITES WITH SPHERULITES.	This work				
	TRAVERTINE & (TUFA)		Banded carbonate rock formed by precipitation of $CaCO_3$ around hot springs (chiefly), by loosing CO_2 to evaporation. The spongy type is called **TUFA**.		Riding (2000); Pettijohn (1957)				
Bound or not bound components	LAMINITE		Fine grain size carbonate rock (muddy and/or peloidal) formed by recurrent thin laminations. Laminations mostly plane-parallel, smooth surfaces (microbial or not) or crenulated (microbial origin).	SMOOTH	Demicco (1994)	Carpet, Microbial carpet, Microbial mat, Algal mat, Microbial laminite, Biolaminite			
				CRENULATED					
Unrecognizable original texture	CRYSTALLINE LIMESTONE		Totally recrystalized carbonate rock. Impossible identify original depositional texture. MICROCRYSTALLINE LIMESTONE: crystal sizes between 5 and 50 µm.		Folk (1962), Dunham (1962)	Sparite Microsparite			
	DOLOMITE		Totally dolomitized carbonate rock. Impossible identify the original depositional texture. MICRODOLOMITE: crystal sizes between 5 an 50 µm.			Dolosparite Microdolosparite			

Table AP1.1 - Carbonate rock classification scheme proposed by Terra et al., (2010)

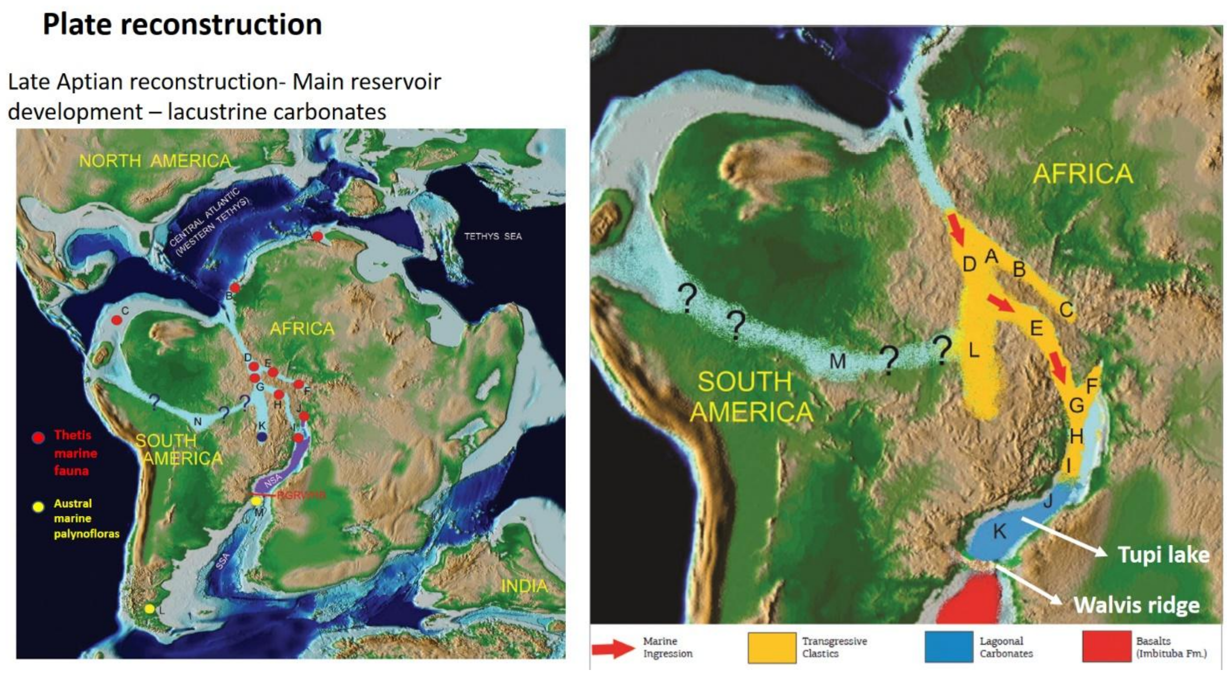

Figure AP1.7- Late Aptian plate reconstruction showing that the Barra Velha carbonates were deposited in a Continental paleo lake (Tupi Lake) which was isolated from Proto South Atlantic Ocean (SSA) by the Walvis Ridge. No marine fossils were found within the Tupi paleo lake. (Modified from Arai, 2014)

The abiotic-biotic dichotomy (2014 – 2019)

Despite the widespread using of the term “microbialite” to refer to the Barra Velha Formation carbonates, no evidence of a microbial induced deposition was found in the previous literature. Wright and Barnett (2015) recognized three main components of the BVE carbonate system: laminated mud carbonates, millimeter-diameter spherulites with Mg- clay matrix, and centimeter-scale shrub-like growths. They proposed that Mg-silicate precipitation took place during evaporation epochs at pluvial controlled paleo lakes. They also advocate that shrub-like carbonates are result of abiotic fascicular calcite crystals precipitation during epochs of low microbial gel precipitation rates (Figure. AP1.8.)

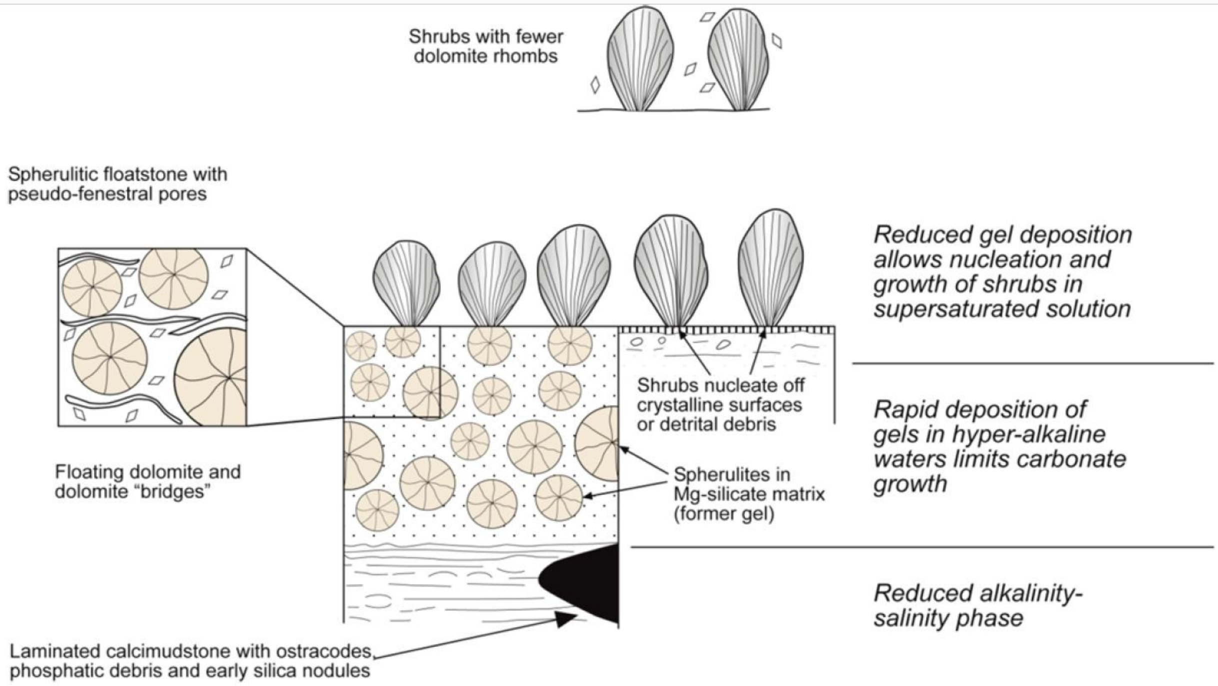

Figure AP1.8- Abiotic induced precipitation model proposed by Wright and Barnett (2015).

After 2014, due to both corporative and government decisions, data from Brazilian pre-salt basins were released for academic research. The main works are those of Souza et al., (2018) and Farias et al (2019). The first one defined that the shrubby carbonates are the results of abiotic precipitation induced by combined evaporation and hydrothermal fluids circulation during the deposition. At the basement horsts perennial highs, the hydrothermal fluids ascended through deep faults and provided cations and CO_2 for fascicular calcite shrubs in reeflike structures called chemically induced build ups. They also carried out an extensive scanning electron microscopy looking for any microbial activity feature, but nothing was found. Together with isotopic data of the same study, the presence of dawsonite ($NaAl(CO3)(OH)_2$) and saddle dolomite indicates a hydrothermal origin of the carbonates with a high pressure of CO_2 in the system. Hydraulic fracturing and brecciation related to hydrothermal circulation were also reported. As a result, they proposed a model of lacustrine environment carbonate precipitation in which both cations and anions were provided by deep hydrothermal conduits. At the vicinity of both subaerial and subaqueous travertines, there are reworked deposits made mainly by grainstones and minor rudstones. At flat platforms they described in situ laminites deposition and associated reworked facies in which, during early diagenesis, spherulites were precipitated. At the lows they described in situ Mg-clay deposition due to evaporation and early diagenesis (Figure. AP1.9).

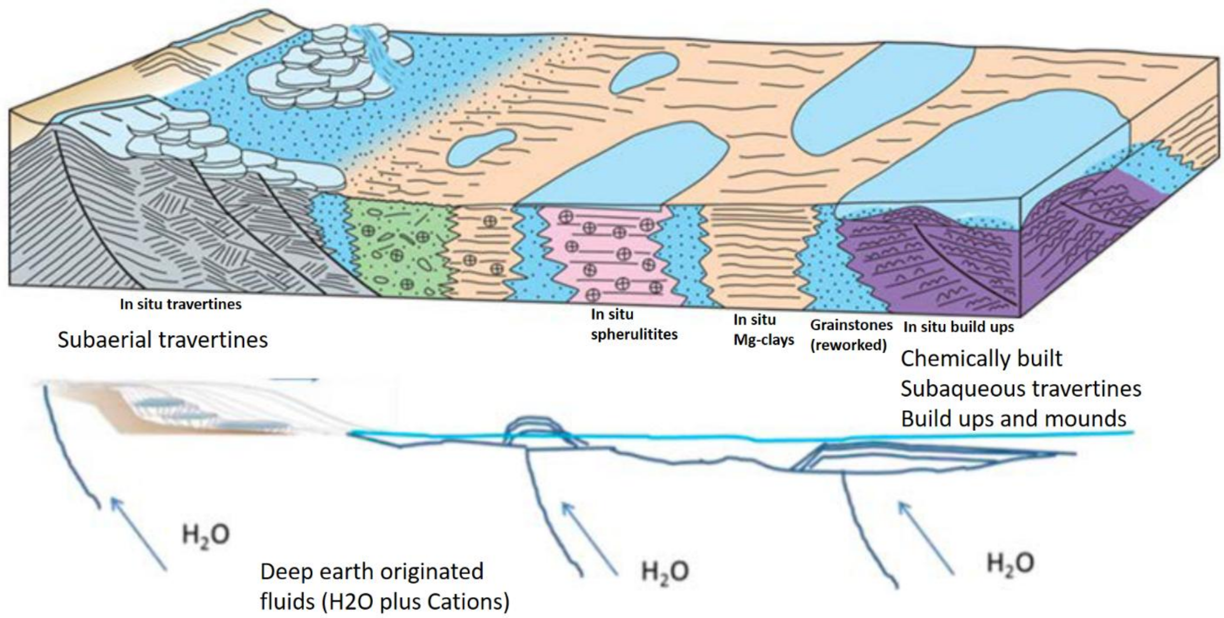

Figure AP1.9- Depositional model proposed by Souza et al., (2018). The deposition of BVE carbonates was dominantly controlled by abiotic factors.

Farias et al., (2019) also proposed an abiotic origin for the carbonates, but with a strong evaporation component. Regarding the Walvis-Rio Grande volcanic high as a topographic barrier during the Aptian, they suggested that the main sources of calcium to the Tupi paleolake were hydrothermal brines, mostly from infiltrating seawater that reacted at depth with the basalts of the volcanic barrier. Hyperextension of the lithosphere, with a shallow Moho, increased geothermal gradients to levels favorable for the generation of hydrothermal $CaCl_2$ brines, a process facilitated by the already elevated Ca++ ion and low sulfate concentration of the Cretaceous Ocean.

The discovery of a regional scale pervasive hydrothermal alteration process (2017-2020)

The first clue of an abiotic component in the pre-salt petroleum system was provided by the research of Santos Neto et al., (2012). They carried out an extensive isotopic analysis of PVT samples from pre-salt reservoirs and found signatures represented by almost exclusively CO_2 originated in the mantle (negative delta C13 values and relatively high values for 3He/4He) (Figure AP1.10).

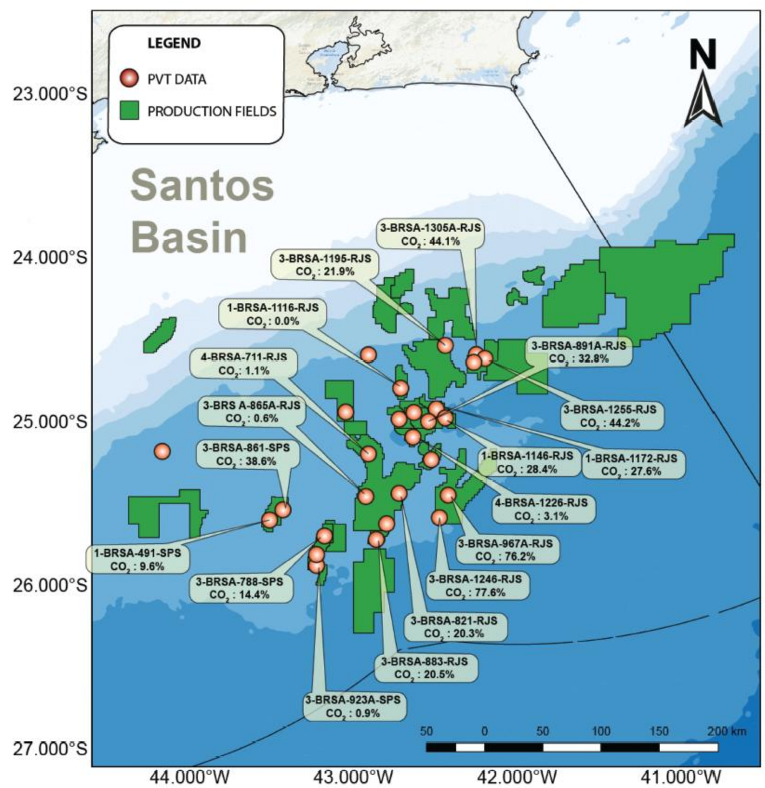

Figure AP1.10- CO2 concentrations in the Santos Basin pre-salt accumulations (Santos Neto et al., 2012).

With the discovery of the Pao de Açucar field in Campos basin, the first indications of hydrothermal processes affecting the Aptian lacustrine carbonate reservoirs were published by Vieira de Luca et al., (2017). They described a pervasive hydrothermal silicification that in some cases obliterated the original fabric. They also indicated that the reservoir quality was enhanced with all overall porosity gain due to dissolution and formation of cave porosity (Figure. AP1.11 and AP1.12).

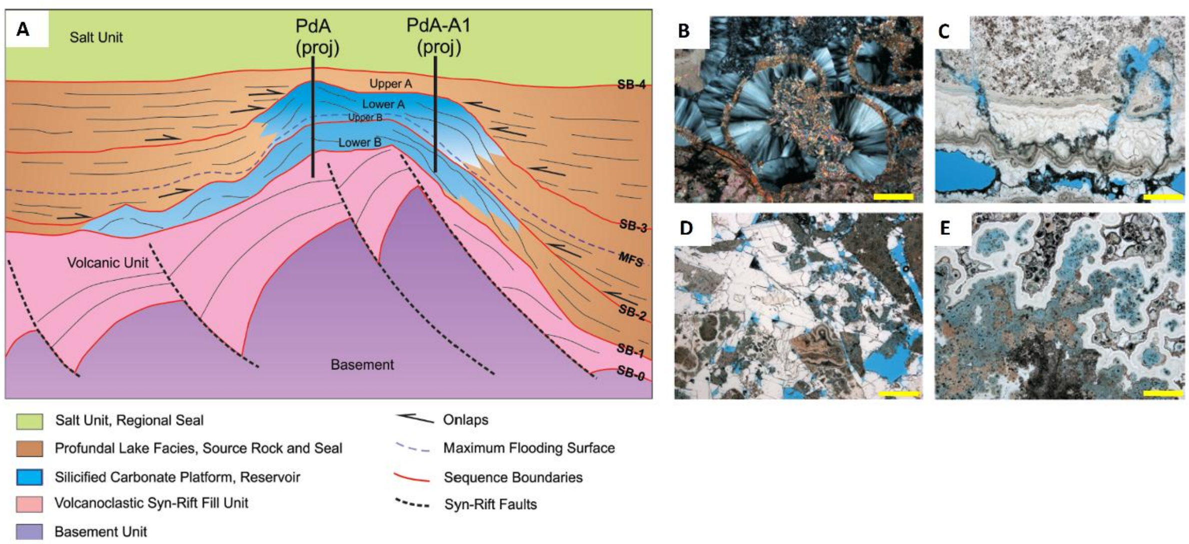

Figure AP1.11- Hydrothermal silicification al the Pao de Açucar Field at Campos Basin. A) Geological section showing the main reservoir units; B) Thin section photo of silicified deep lake spherulitites; C) Silicified hydraulic vein; D) Late dolomitization of a hydraulic breccia; E) Heavily silicified microbial carbonate with opal vugs; in all thin section images the scale bar is 200 micrometers. (Modified from Vieira de Luca et al., 2017).

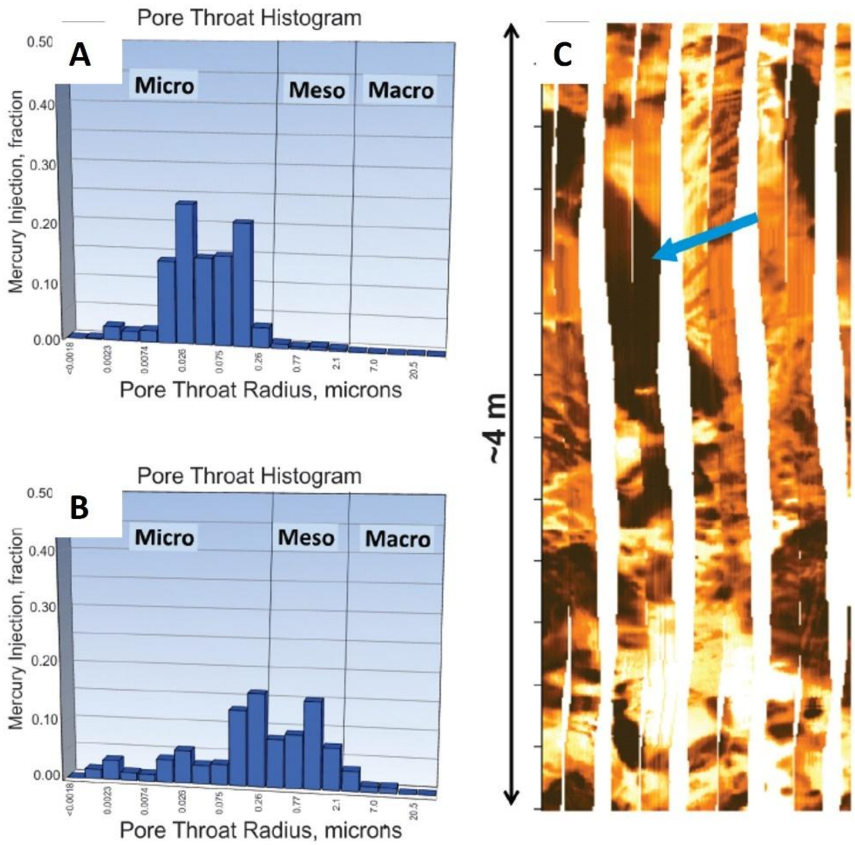

Figure AP1.12- Pore size distribution after hydrothermal silicification at the Pao de Açucar carbonates. The pore throat sizes were augmented at the (A) micro, (B) meso and (C) Giga-pore (crystal caves) scales at image log. (Modified from Vieira de Luca et al., 2017).

At the South Atlantic conjugate margin of Angola, at Kwanza basin, Poros et al., (2017) reported the same type of hydrothermal alteration using petrography, fluid inclusion, and raman spectroscopy. They proposed a hydrothermal activity related to deep faults trough which silica, CO_2 and CH_4 rich fluids ascended to the sedimentary pile and led to the overall porosity increment (Figure. AP1.13).

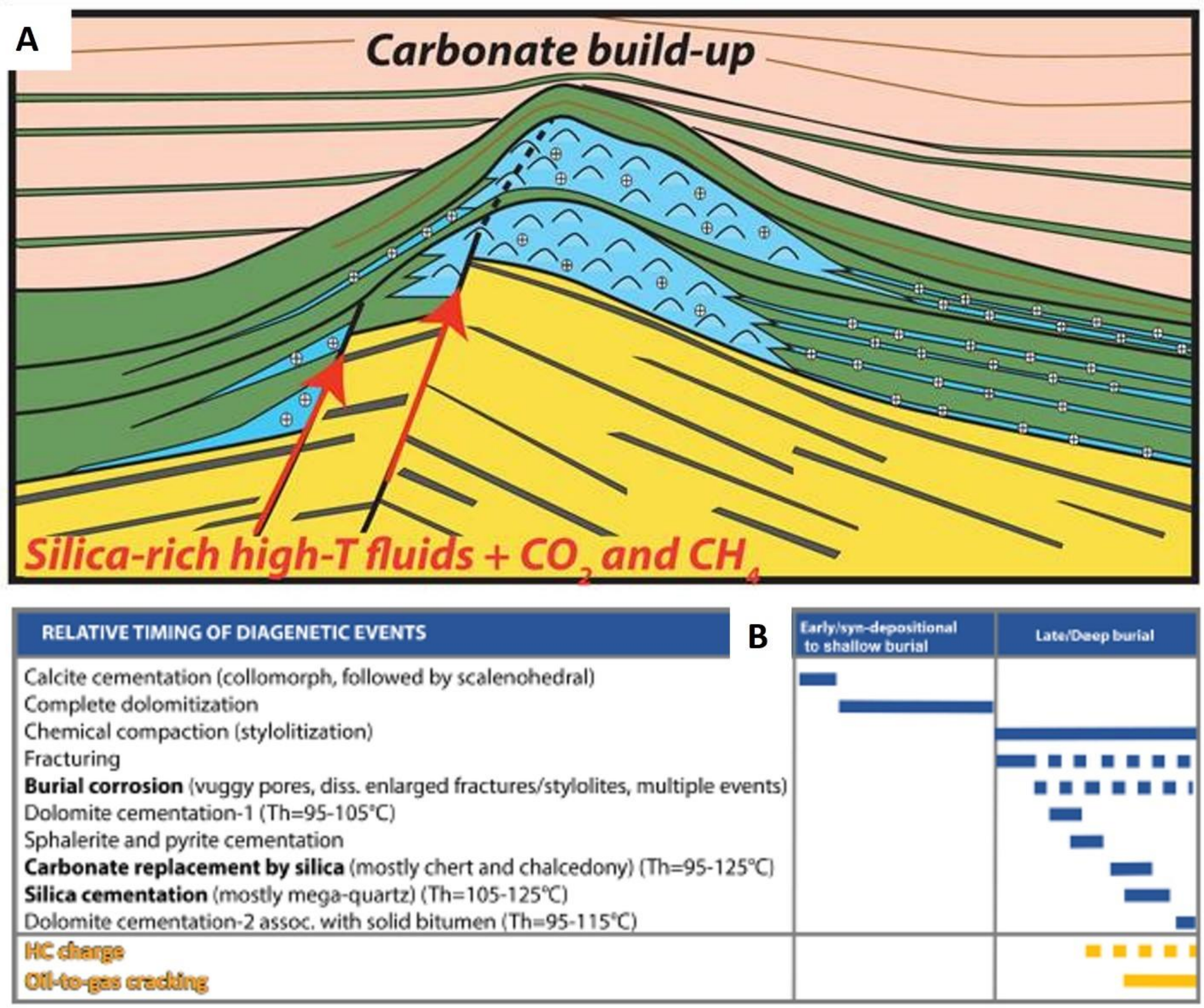

Figure AP1.13- A) Hydrothermal silicification model and B) Diagenetic history proposed by Poros et al., (2017) for the Kwanza basin Aptian lacustrine carbonates.

In terms of paleohydrology, Pietzsch et al., 2018 integrated strontium isotopes and biostratigraphic data to constrain the hydrodynamics during the deposition of the BVE carbonates at Santos Basin. They proposed a combination of downward meteoric and upward crustal fluid mixing in a high geothermal gradient setting due to crustal stretching. The meteoric waters were responsible for the aquifer feeding and the high geothermal gradient was the heat source which has driven the basin scale fluids circulation.

Coming back to the Brazilian side, Lima et al., (2020) carried out an extensive study on the origin of the hydrothermal alteration of the BVE chrono correlates lacustrine reservoirs at northern Campos Basin. They integrated systematic petrography, cathodoluminescence, scanning electron microscopy, microprobe, x-ray diffraction, fluid inclusion, and isotopic studies in rock samples from exploratory wells drilled by Petrobras at that oil province. Their main discoveries were that the deep burial of the lacustrine carbonates combined with ascending silica rich fluids promoted a pervasive hydrothermal alteration characterized mainly by dolomitization, silicification, and dissolution at varying degrees and intensities (Figure. AP1.14 and AP1.15). They also proposed that the main controls on the creation, redistribution, and obliteration of porosity in the pre-salt reservoirs are the eo-diagenetic minerals precipitation and dissolution due to variations in the lake water chemistry and the episodic flow of hydrothermal fluids under burial conditions. According to these authors, the hydrothermal processes also controlled the anhydrite precipitation at the very top of the lacustrine carbonate reservoirs (Figure AP1.16).

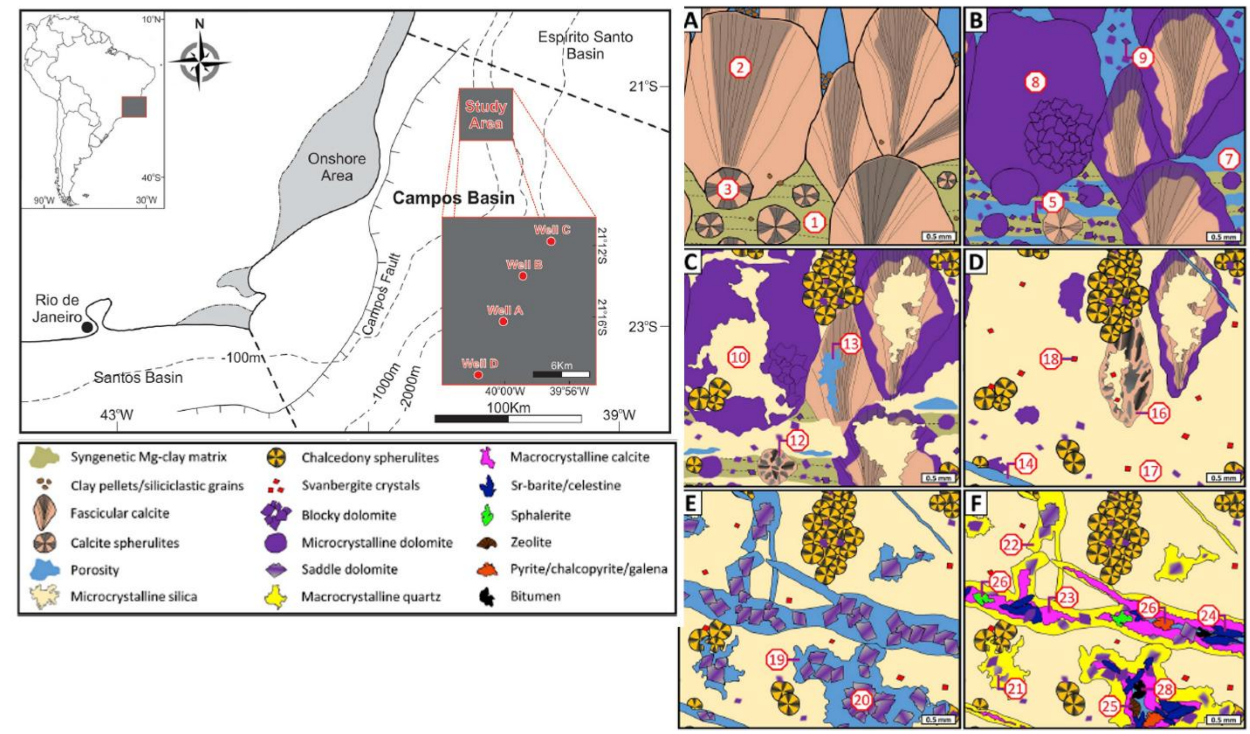

Figure AP1.14- Hydrothermal events sequence at northern Campos Basin lacustrine carbonate reservoirs. (A) Syngenetic crust of coalesced fascicular calcite aggregates covering syngenetic Mg-clay matrix. B) Microcrystalline and blocky dolomite partially replacing the Mg-clay matrix, calcite spherulites, and fascicular aggregates. C) Microcrystalline quartz and chalcedony spherulites partially to fully replacing preexisting constituents, and filling primary and secondary porosity. D) Silicification of some dolomitized fascicular and spherulitic aggregates. E) Intense silicification, fracturing and dissolution. Saddle dolomite partially filling fracture and vugular pores. F) Macrocrystalline calcite, radiated Sr-barite and celestine, fibro-radiated zeolite, euhedral to subhedral pyrite, chalcopyrite, galena and sphalerite, and massive bitumen filling the secondary pores. The numbering is in accordance with the paragenetic sequence in Figure. AP1.15.

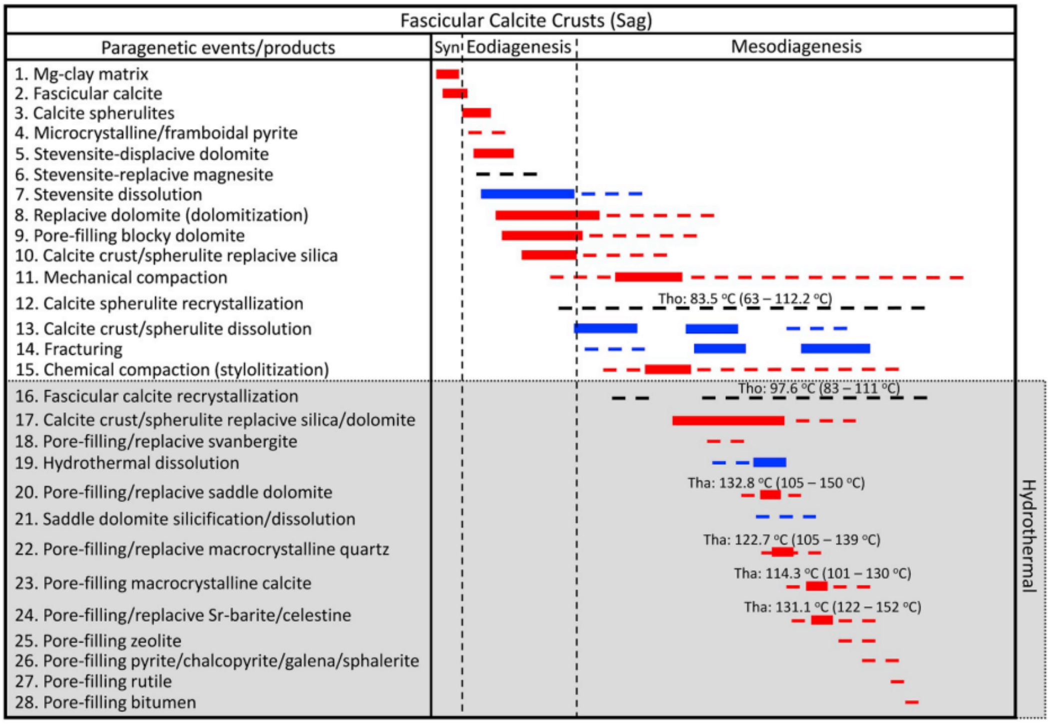

Figure AP1.15- Sequence of diagenetic and hydrothermal events affecting the lacustrine carbonate reservoirs at northern Campos Basin. (Lima et al., 2020)

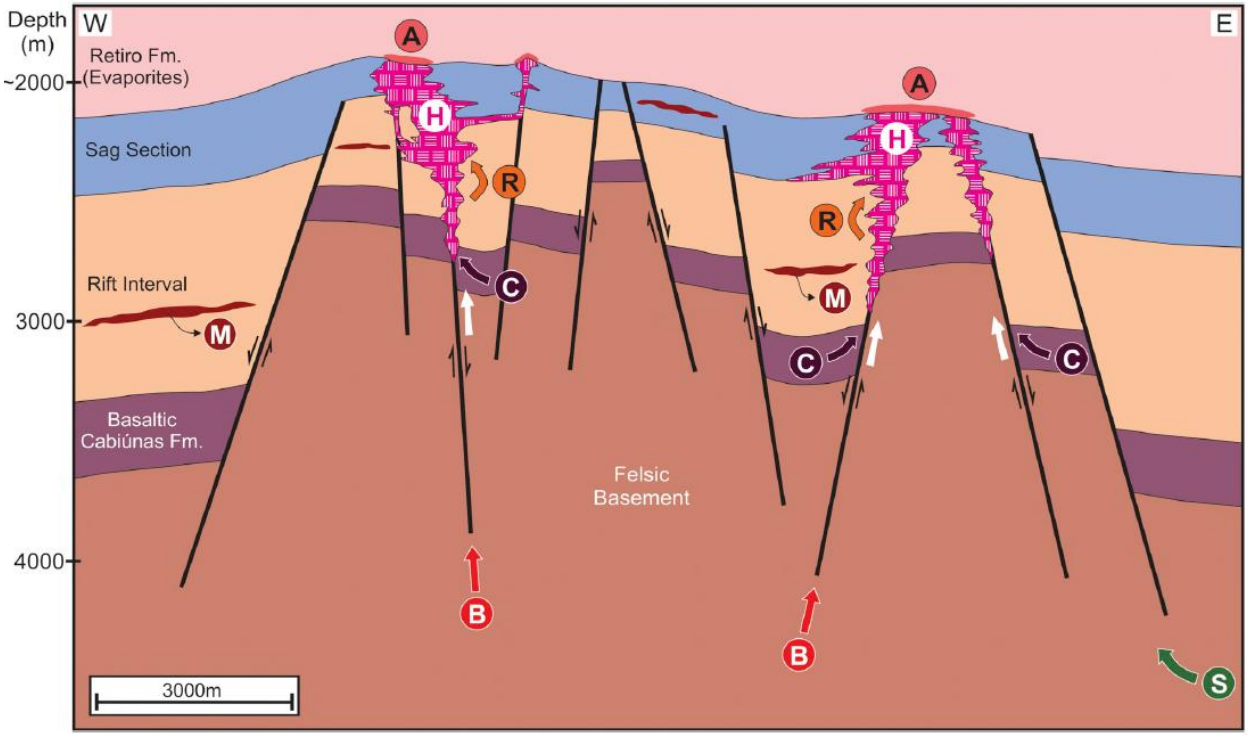

Figure AP1.16- Proposed model (Lima et al.,2020) for the origin of the hydrothermal alteration of the Aptian lacustrine carbonate reservoirs at northern Campos Basin. The fault-focused hydrothermal system probably involved mixing of fluids derived from several sources: (B) Pre-Cambrian basement (mainly granitic-gneissic felsic rocks), (C) volcanic rocks from the Cabiúnas Formation (mainly basaltic), (R) rift deposits from the Atafona and Coqueiros formations, and (S) serpentinization of the upper mantle. Note the intrusive magmatic (mafic) rocks (M) in the Coqueiros (rift interval) and Macabu (sag section) formations; hydrothermal alteration (H) comprising extensive dolomitization, silicification, and dissolution, with the paragenesis including saddle dolomite, macrocrystalline calcite, mega-quartz, Sr-barite, celestine, fluorite, dickite, sphalerite, galena, and other metallic sulfides filling fractures and dissolution porosity; and anhydrite thick layer (A) at the top of the sag reservoir affected by the hydrothermal alteration (base of evaporites).

Towards the unification theory and the public pre-Salt public data avaliability "spring" (2020 – 2024)

With an aim to standardize the petrographic classification, Gomes et al., (2020) proposed a simplistic new scheme with three end members for both in situ and reworked facies (Figure. AP1.17). For the in-situ facies, they proposed a ternary diagram with the following end members:

- Shrubs: they are composed of fibrous to extremely coarse bladed calcite crystals. Some of them have incorporated random fine detrital grains, or rare bioclasts. At individual layers, shrubs are parallel with each other, ranging from densely packed (0.2 mm spaced) to widely spaced (4 mm). They occur dominantly as horizontal layers and locally form dome shaped structures (build ups).

- Spherulite: spherical to sub-spherical calcite aggregates with radial extinction and often dolomitized and recrystallized. Their diameters range from 0.1 to 4 mm (mean= 1 mm). When the recognition is possible, their nuclei are made of clay minerals, micrite and rare ostracod fragments. When silicified, the calcite is replaced by micro-quartz.
- Mud: It is a fine-grained matrix composed mainly of clay minerals, calcite, dolomite and silica.

For the reworked facies they adopted another ternary diagram for grains composition endmembers:

- Shrub grains
- Spherulite grains
- Mud grains

They also addressed the fine-grained carbonates issue (mainly composed of Mg-silicates) with four end members in terms of mud composition:

- Mg-clays: mainly aluminum free Mg-silicate minerals formed and preserved in the sub-littoral zones. They can precipitate under both biotic and abiotic influence, depending on local bacterial distribution and evaporation rates. They suggest low energy, alkaline lacustrine environments.

- Calcite/dolomite: both microcrystalline calcite and/or dolomite have clearly replaced Mf-clay minerals. They often occur together suggesting replacement of calcite by dolomite. Both of them might have formed during either deposition or diagenetic stages.
- Silica: microcrystalline silica occurs in variable manners: as hydrothermal deposits, silica nodules and chert layers alternating with mudstones. In this last case it might be related to the dissolution of Mg-clay minerals.

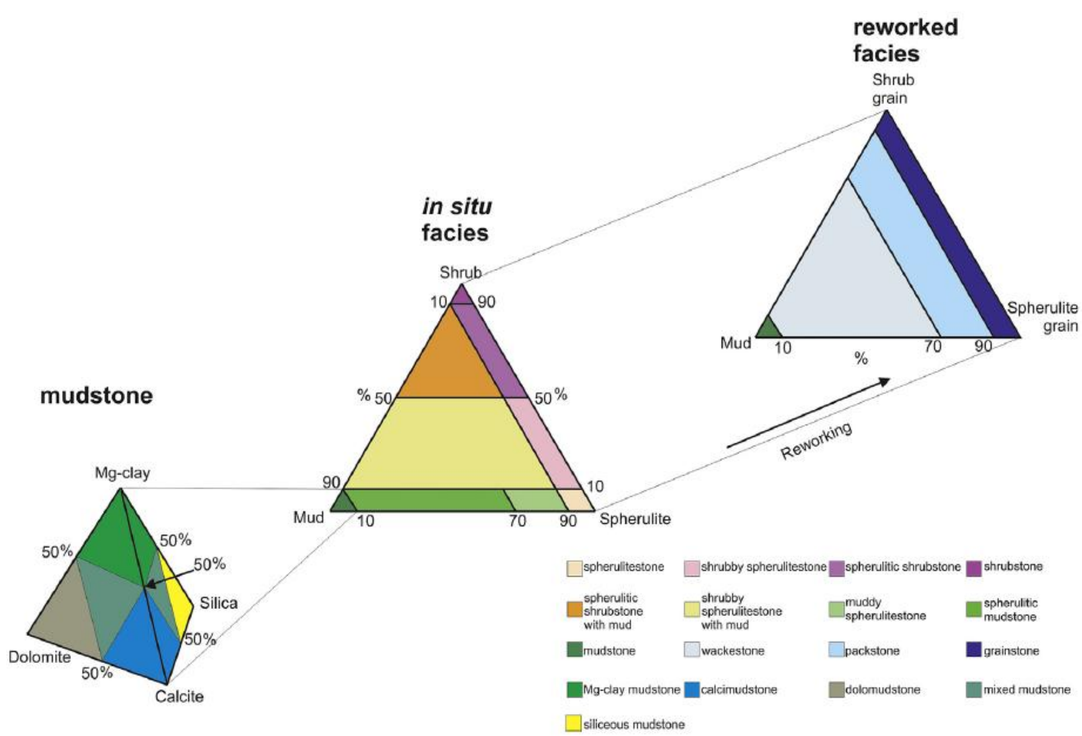

Figure AP1.17- Petrographic classification scheme proposed by Gomes et al., (2021) for the Barra Velha lacustrine carbonates.

Sartorato et al., (2018) when working with facies modeling at the Mero Field, proposed a mixed model for the lacustrine carbonate reservoirs of the Barra Velha. They described hydrothermally induced shrubs build ups formed due to deep crustal fluids circulation through rift faults (subaqueous travertines) and at their vicinity they are reworked mainly as shrub grains grainstones intercalated with crenulated laminites (microbialites) (Figure. AP1.18).

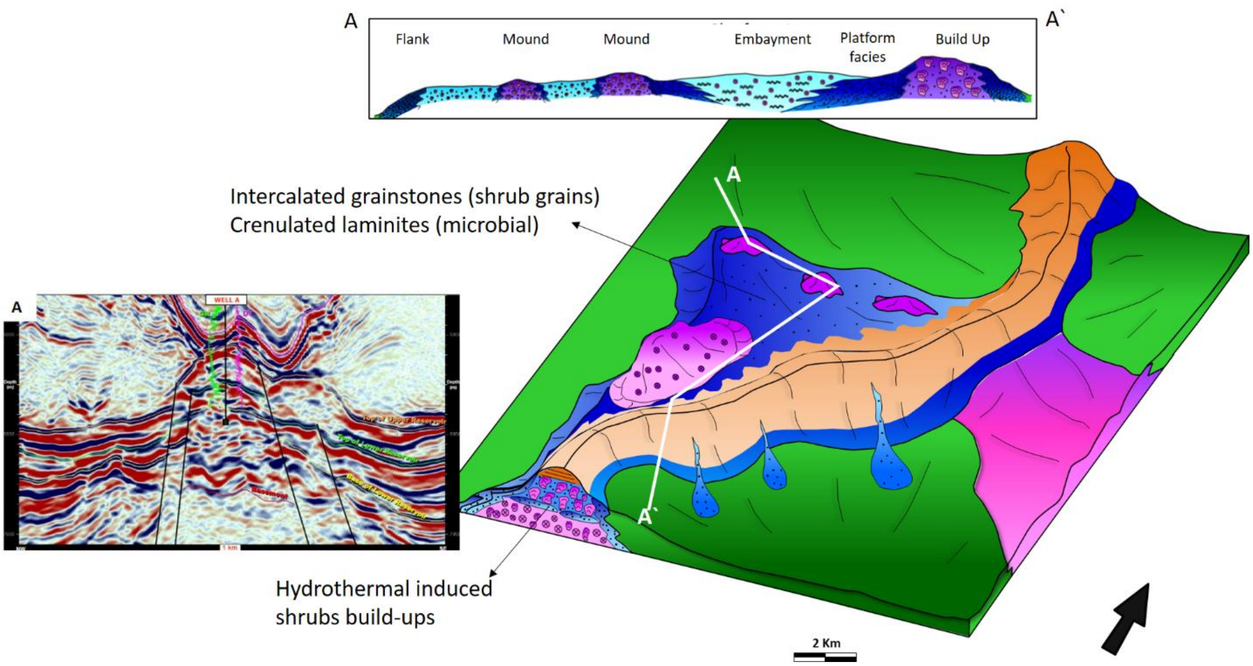

Figure AP1.18- Depositional model proposed for the BVE carbonates at Mero Field based on well and seismic data (modified from Sartorato et al., 2018).

During the last five years (2020-2025) a lot of data on pre-salt oilfields and even dry wells are available for academic research (public data). In this sense, several research papers on geology, geophysics and engineering were published. Finally, in 2024, Petrobras released its e-book on pre-salt carbonate reservoir geosciences, where among several chapters, Rancan and Castro (2024) and Rancan et al., 2020, also showed the evolution of the geological knowledge of the BVE carbonate reservoirs.

The relentless search for analogs

Since their discovery in 2006, all petroleum companies operating in both conjugate margins (mainly Santos and Kwanza) tried to find an appropriate analog for these unique carbonates. After more than a decade, the best analogs were found at the Basin and Range province (USA) and the Afar Region (East Africa). In the USA, geoscientists studied outcrops at Green River, Lake Idaho, Great Salt Lake, Pyramid Lake, Mono lake and Winnemuca Lake. In the last example, Virgone et al., (2021) reported both spring and microbial controlled lacustrine carbonate deposits.

At the Afar triangle, the best analogs are described by Awaleh et al. (2015) at the Lake Abhe (Danakil depression, Djibouti), where travertine towers are observed aligned over pre-existing faults (Figure. AP1.19). Cavallazzi et al., (2019) described hydrothermal salt deposition in extreme environments at the geothermal Dallol area (Ethiopia).

Figure AP1.19- Travertine towers at the Lake Abhe, Djibouti (Awaleh et al., 2015).

Cupertino (2023) studied in detail the Green River Fm (Eocene, Basin and Range province, USA) carbonates and consider them as important outcrop analogues for the Barra Velha formation reservoirs (Figure AP1.20).

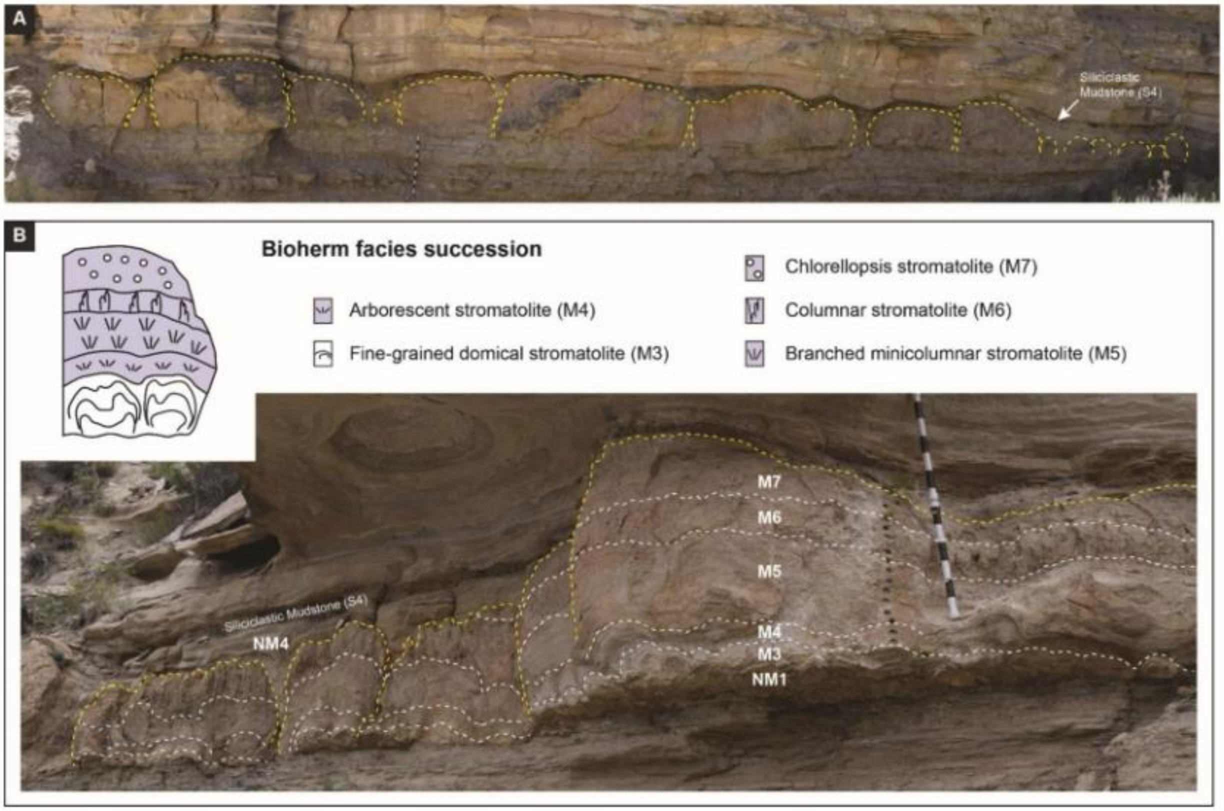

Figure AP1.20 – Outcrop study of Eocene carbonates from Green River Fm (Cupertino, 2023). Bioherm facies succession: A) Individuals bioherms, west of Rosewood Section, showing distinctive domical shape. B) distinctive facies succession within bioherms: fine-grained domical stromatolite facies (M3), followed by arborescent (M4) and branched minicolumnar stromatolites (M5), columnar stromatolites (M6), and Chlorellopsis (M7) facies at the top. Non-microbialite facies: Intraclastic conglomerate facies (NM1) and Ostracodal grainstone facies (NM4).

Discussion

The scientific knowledge evolution on the Barra Velha carbonates followed a normal trend. As stromatolite-like structures were first observed, it is normal and acceptable to link them to microbial activity, as most of the known examples are linked to. In my opinion, the Dunham classification scheme is good enough to describe these carbonate reservoirs. For instance, for the fascicular calcite aggregates shrubs, the term shrubby boundstone should work. In terms of matrix composition, it seems that the fact of micrite rareness is not a big issue, since the Mg-clay has the same rule in terms of reservoir quality and paleoenvironment.

On the biotic-abiotic dichotomy about the origin of carbonates, I think that there are no end members in nature. Microbes are everywhere, even around hot springs (Cavallazzi et al.,2019), so biotic and abiotic processes coexist in time and space. Sartorato et al. (2018) showed a good example in the Mero Field, there we have both hydrothermal travertines made by shrubby fascicular calcite aggregates forming build up ridges and crenulated laminites at the neighboring platform.

The regional hydrothermal silicification seems to be a process related to crustal hyper stretching during both Early Aptian rift stage and Late Aptian sag phase (lacustrine carbonates deposition). The mantle degassing process which is responsible for the CO_2 availability for either carbonate precipitation or hydrothermal fluids circulation was reported in other tectonic settings.

Pinto (2014) performed detailed research on the linkage of tectonic evolution and fluid history in hyperextended rifted margins. One of his main findings after conducting field observations, structural analysis, mineralogical studies, geochemical and geophysical data analysis, was the demonstration of the existence of a regional fossil fluid circulation record at both Alpine and Pyrenean rift systems (Figure. AP1. 21)

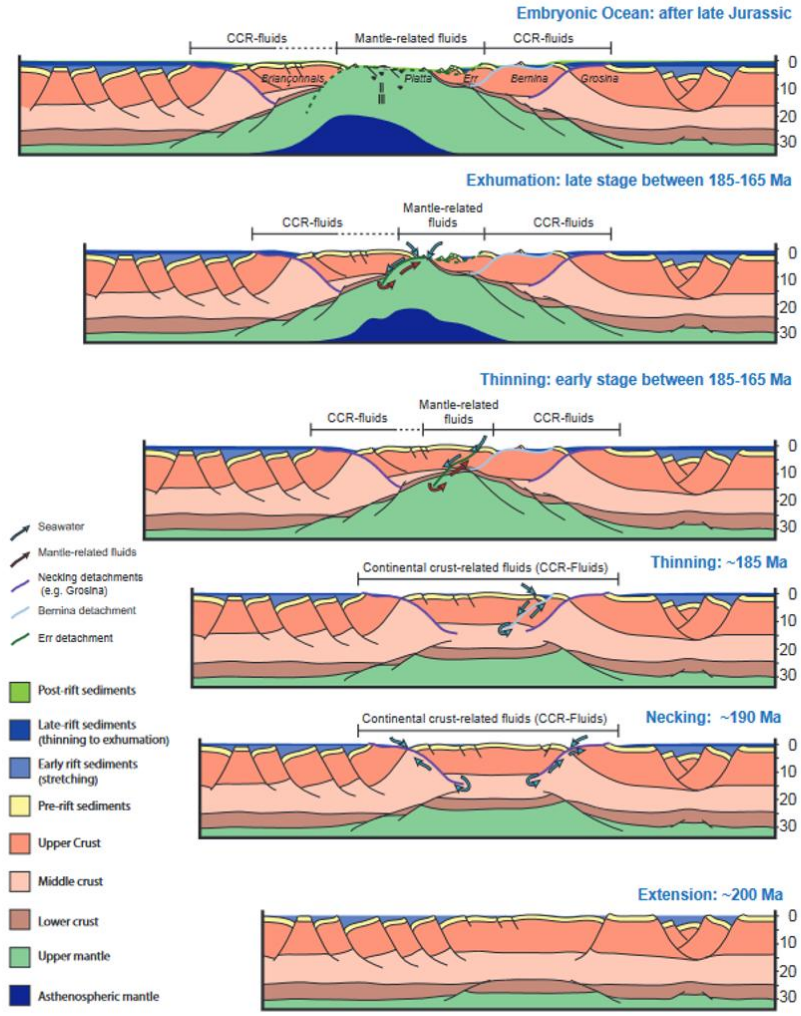

Figure AP1.21- The link between tectonic evolution and fluid migration along the Alpine Tethyan margin. The Grosina and Bernina detachments record a fluid with a continental composition (about 190-185 Ma). Between 185 and 165 Ma, the ED fault rocks and the supradetachment basin record mantle-related fluids that are produced by initial serpentinization under thinned continental crust. Serpentinization and fluid flow evolve with the progressive mantle exhumation toward the seafloor. (Extracted from Pinto,2014).

This tectonically influenced crustal scale fluids circulation was also observed by Virgone et al., (2021). They described the main favorable geodynamic domains for continental carbonates deposition:

a) both poor and magma-rich rifted margins

b) orogens.

c) Foreland and retro-foreland basins

At each one of the systems, three types of CO_2 sources are competing: magmatic, atmospheric and biosphere sourced.

For the Santos-Kwanza hyperextension rift initiated passive margin, Araujo et al. (2022) performed a numerical modelling of the lithospheric stretching. They showed that during both Early Aptian rift and Late Aptian sag, the continental crust was as thin as 5-10 km and the 500ºC geotherm was very shallow. This combination led to continuous mantle degassing (mainly CO_2 releasing) and both deep and meteoric hydrothermal fluids circulation. These processes influenced the carbonate deposition and hydrothermal alteration (Figure. AP1.22).

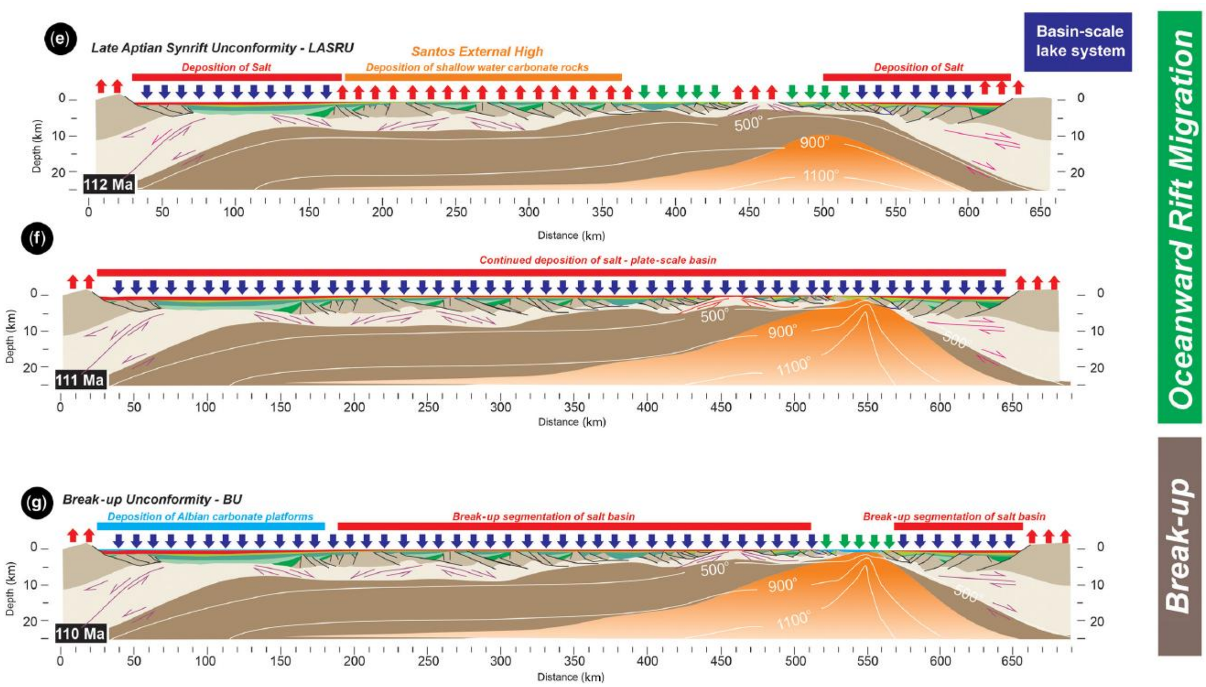

Figure AP1.22 - Numerical modeling results of Aptian evolution of Santos-Kwanza conjugate margins. (Araujo et al., 2022)

Conclusions

- The Pre-Salt petroleum system is in a hyper-stretched continental crust domain in which continuous mantle degassing contributed to both carbonate precipitation (abiotic and microbial) and other mantle and crustal fluids circulation.
- Both petrography and petrogenesis started with a pure microbial model and evolved for abiotic and mixed model, with a unifying petrographic classification scheme.
- Mantle degassing is also responsible for CO_2 content in the petroleum that ranges from traces up to >99% (In Petrobras' oil fields it is separated and reinjected for enhanced oil recovery purposes).
- The other fluids (mainly silica rich) are probably a mix of several origins (mantle + crust + sediments + meteoric) that caused a regional and pervasive hydrothermal silicification of the Aptian lacustrine carbonates at both conjugate rifted margins (Brazil and Angola).
- These hydrothermal fluids are also being considered as a source of both CO_3 and anions for abiotic fascicular calcite precipitation (Shrubby build ups) – the best reservoirs of all.
- This reservoir silicification can be weak to severe, with total destruction of original fabric. But in the mounds, it led to a net porosity increase (hydrothermal

corrosion of pre-existing karsts and eo-diagenetic dolomites). Highly silicified reservoirs are excellent producers (one vertical well = 40,000 – 60,000 bpd).

- Analogs to these reservoirs occur in the Basin and Range (USA) and Afar triangle (Ethiopia) – regions of high geothermal gradient and crustal stretching.

References

Arai, M. 2014. Aptian/Albian (Early Cretaceous) paleogeography of the South Atlantic: a paleontological perspective. Brazilian Journal of Geology. 44. 339-350. 10.5327/Z2317-4889201400020012.

Araujo, M., Perez-Gussinye, M., Muldashev, I. 2022. Oceanward rift migration during formation of Santos-Benguela ultra-wide rifted margin. Geological Society, London, Special Publications. 524. SP524-2021. 10.1144/SP524-2021-123.

Awaleh, M., Bouraleh, F. Boschetti, T., Soubaneh, Y., Moussa, N., Elmi, S., Mohamed, J., Khaireh, M. 2015. The geothermal resources of the Republic of Djibouti — II: Geochemical study of the Lake Abhe geothermal field. Journal of Geochemical Exploration. 159. 129-147. 10.1016/j.gexplo.2015.08.011.

Carminatti, M.S., Cunha, R. A., Moraes, M.A.S., Matsunaga, V.Y. 2019. Brazil: An Update on the Campos and Santos Super Basins, and recent discoveries. AAPG conference and Exhibition. San Antonio.

Carvalho, M. D., and F. L. Fernandes, 2021, Pre-salt depositional system: Sedimentology, diagenesis, and reservoir quality of the Barra Velha Formation, as a result of the Santos Basin tectono-stratigraphic development, *in* Marcio R. Mello, Pinar O. Yilmaz, and Barry J. Katz, eds., The supergiant Lower Cretaceous pre-salt petroleum systems of the Santos Basin, Brazil: AAPG Memoir 124, p. 121–154.

Cavalazzi, Barbara & Barbieri, Roberto & Gó, F & Capaccioni, B & Olsson-francis, Karen & Pondrelli, Monica & Rossi, Angelo Pio & Hickman-Lewis, Keyron & Agangi, A & Gasparotto, G & Glamoclija, Mihaela & Ori, Gian-Gabriele & Rodríguez, Nuria & Hagos, Miruts. (2019). The Dallol Geothermal Area, Northern Afar (Ethiopia)-An Exceptional Planetary Field Analog on Earth. Astrobiology. 19. 10.1089/ast.2018.1926.

Dias, J. L. Tectônica, Estratigrafia e Sedimentação no Andar Aptiano da margem leste brasileira. 2005. Boletim de Geociências da Petrobrás, Rio de Janeiro v. 13, n. 1, p.7-25.

Dunham, R.J., 1962. Classification of carbonate rocks according to depositional texture. In: In: HAM, W.E. (Ed.), Classification of Carbonate Rocks, vol. 1. American Association of Petroleum Geologists, Memoir, Tulsa, pp. 108–122.

Farias, F., Szatmari, P., Bahniuk, A., Franca, A.B., 2019. Evaporitic carbonates in the presalt of Santos Basin – genesis and tectonic implications. Mar. Pet. Geol. 105, 251–272.

Grotzinger, J. P.; James, N. P. Precambrian Carbonates. **Society of Economic Paleontologists and Mineralogists**, Special Publications, 1999.

Lima, B. E., Tedeschi, L., Pestilho, A. L., Santos, R., Vazquez, J., Guzzo, J., De Ros, L. 2020. Deep-burial hydrothermal alteration of the Pre-Salt carbonate reservoirs from northern Campos Basin, offshore Brazil: Evidence from petrography, fluid inclusions, Sr, C and O isotopes. Marine and Petroleum Geology. 104143. 10.1016/j.marpetgeo.2019.104143.

Mimoun, J.G., Fernández-Ibáñez, F. 2023. Carbonate excess permeability in pressure transient analysis: A catalog of diagnostic signatures from the Brazil Pre-Salt, Journal of Petroleum Science and Engineering, Volume 220, Part A, 111173, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2022.111173.

Moreira, J.L.P., Madeira, C.V., Gil, J.A., Machado, M.A.P. 2007. Bacia de Santos. Rio de Janeiro. Boletim de Geociências da Petrobras, v. 15, n. 2, 531-549.

Pietzsch, Raphael & M Oliveira, Daniel & Tedeschi, Leonardo & Neto, Joao & Figureueiredo, Milene & Vazquez, Joselito & Souza, Rogério. (2018). Palaeohydrology of the Lower Cretaceous pre-salt lacustrine system, from rift to post-rift phase, Santos Basin, Brazil. Palaeogeography, Palaeoclimatology, Palaeoecology. 507. 10.1016/j.palaeo.2018.06.043.

Poros, Z., Jagniecki, E., Luczaj, J.,Kenter, J., Gál, B., Correa, T., Diniz Ferreira, E. L., Heumann, M., Elifritz, A., Johnston, M., Mcfadden, K., Matt, V. 2017. Origin of Silica in PreSalt Carbonates, Kwanza Basin, Angola. 10.13140/RG.2.2.31863.42405.

Rancan, C.C.; Candido, A.G.; Quintaes, C.M.S.P.; Castro, M.R.; Oliveira, C.M.M.; Dias, K.D.N.; Sampaio, R.F. 2022. From the arcadian to the contemporaneous: the curios history of pre-salt's section knowledge evolution in the Campos Basin. In: Brazilian Petroleum Conference, 2022, Rio de Janeiro. ABGP, p. 27.

Santos Neto, E., Cerqueira, J.R. and Prinzhofer, A. [2012] Origin of CO2 in Brazilian Basins. Proceedings of the AAPG Annual Convention and Exhibition, Long Beach, California, April 22-25.

Sartorato, A.C.L, Mizuno, T.A., Paegle, J.S., Pozzi, R.P.S, Zanatta, A.S., Rosseto, J.A., Oliveira, L.C., Penna, R.M., 2018. Modelagem geológica do reservatório do Campo de Mero. XI Seminário de Interpretação Exploratória. PETROBRAS. Rio de Janeiro.

Sartorato, A.C.L., Tonietto, S.N., Pereira, E., 2020. Silicification and Dissolution Features in the Brazilian Presalt Barra Velha Formation: Impacts in the Reservoir Quality and Insights for 3D Geological Modeling. Rio Oil & Gas, Rio de Janeiro, Brazil. https://doi.org/10.48072/2525-7579.rog.2020.068.

Stica, J. M.; Zalan, P. V.; Ferrari, A. L., 2014. The evolution of rifting on the volcanic margin of the Pelotas Basin and the contextualization of the Paraná-Etendeka LIP in the separation of Gondwana in the South Atlantic. *Marine and Petroleum Geology*, [Amsterdam], v. 50, p. 21.

Terra, G. J. S.; Spadini, A. R.; França, A. B.; Sombra, C. L.; Zambonato, E.; Juschacs, L. C. S. da; Arineti, L. M.; Erthal, M. M.; Blauth, M.; Franco, M. P.; Matsuda, N. S.; Carramal, N. G. da S.; Junior, P. A. M.; D'ávila, R. S. F.; Souza, R. S. de; Tonietto, S. N.; Anjos, S. M. C. dos; Campinho, V. S.; Winter, W. R. 2010. Classificação de rochas carbonáticas aplicável a bacias brasileiras. Boletim de Geociências da Petrobrás, Rio de Janeiro, v.18 (1), p. 9-29.

Viera de Luca, P.H., Matias, H., Carballo, J. 2017. Breaking barriers and paradigms in Pre-Salt exploration: the Pão de Açúcar discovery, (offshore Brazil). AAPG Memoir, 113, 177-194.

Virgone, A., E. Gaucher, P.-A. Teboul, C. Kolodka, and E. Poli, 2021, A discussion on South Atlantic pre-salt continental carbonates: A mosaic of sedimentary facies and processes, *in* Marcio R. Mello, Pinar O. Yilmaz, and Barry J. Katz, eds., The supergiant Lower Cretaceous pre-salt petroleum systems of the Santos Basin, Brazil: AAPG Memoir 124, p. 193–214.

Pinto, V.H. 2014. Linking tectonic evolution with fluid history in hyperextended rifted margins: examples from the fossil Alpine and Pyrenean rift systems, and the present-day Iberia rifted margin. PhD thesis. Universite de Strasbourg. 256p.

Vieira de Luca, Pedro Henrique, Hugo Matias, José Carballo, Jose Luis Algibez Alonso, Jordi Tritlla, Mateu Esteban, et al., 2017, Breaking barriers and paradigms in presalt exploration: The Pão de Açúcar discovery (offshore Brazil), *in* R. K. Merrill and C. A. Sternbach, eds., Giant fields of the decade 2000–2010: AAPG Memoir 113, p. 177–194.

Wright, V. P.; Barnett, A. J. 2015. An abiotic model for the development of textures in some early South Atlantic Cretaceous carbonates. **Geological Society**, London, Special Publication, v. 418, p. 209-219.

Zalan, P. V. ,2014. Comparisons between the Deep Crustal Structure of Magma-Poor and Volcanic Passive Margins. In: Atlantic Conjugate Margins Conference, 4. St. John's, Canada. *Extended Abstracts*... St. John's, Canada: [S.n.], 2014. p.75–78.

Zalan, P. V. 2024. Remaining Potential Brazil DERBYANA 2024. Derbyana. 45. 60 pages. 10.14295/derb.v45.826.

Appendix 2 – Supervised regression toy examples

Decision trees

Classification and regression trees were first implemented by Morgan and Sonquist (1963) and consist of very simple and powerful predictive models building tools used in almost all areas of knowledge. They are the building blocks of random forests that are used for both classification and regression tasks (Ho, 1995).

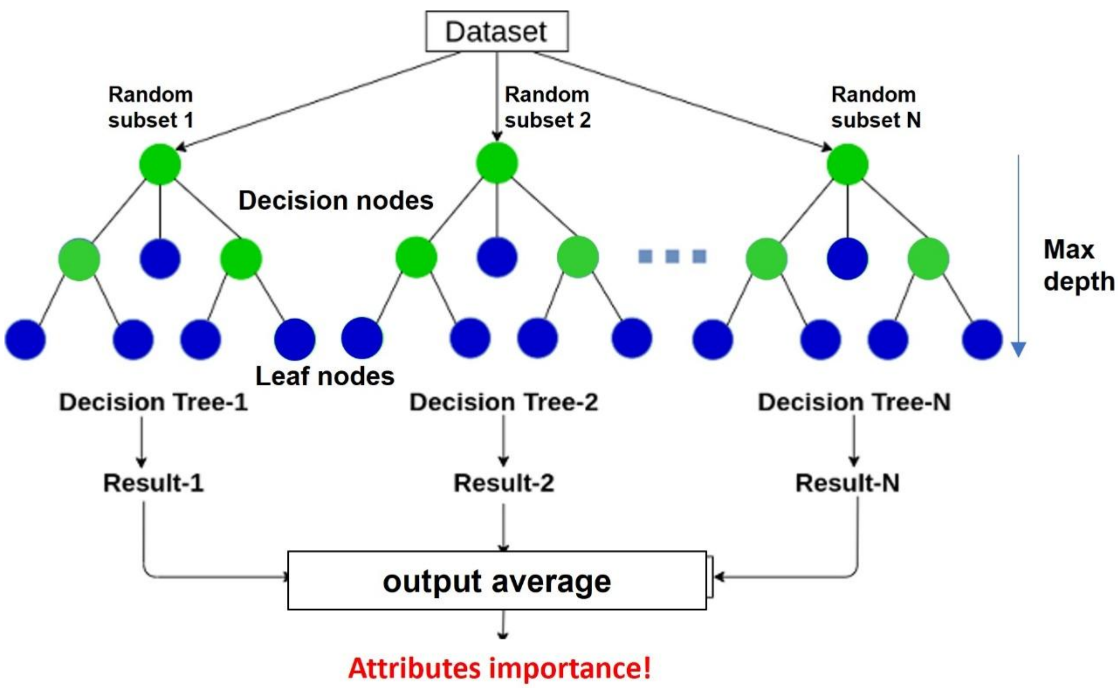

Figure AP2.1 – Random Forest composed of multiples decision trees; the maximum depth is the one composed of only leaf nodes. The final prediction is done by averaging the leaf nodes of each decision tree. Very shallow trees tend to underfit whereas too deep trees tend to overfit the data (modified from Goyal, 2022).

Herein, we present how one simple decision tree was built to fit an also simple, nonlinear, and low dimension (2D) function (y = sin (x) + 20% random noise). Our independent variable dataset (x) is made of 5000 datapoints of values ranging from 0 to 10 with a step of 0.002. The target (y=sin(x)) is calculated after plugging each one of the x values to the equation and adding 20% of random noise, the result is the noisy sinusoidal function shown in the figures below as grey dots. We started fitting a max depth=1 decision tree and then calculated both R^2 and mean absolute error (MAE) between the predicted and actual values of y (Figure AP2.2). Next, we did the same for decision trees with max depths of 2, 3, 5 and 10 (Figures AP2.3 to AP2.6). As can be seen in Figure AP2.3, the best fit was obtained with max depth = 5 (R^2=0.90, MAE=0.17) decision tree. A max depth=1 decision tree is underfitted (R^2=0.21, MAE=0.50), whereas a max depth=10 decision tree, despite its good mathematical fit (R^2=0.90, MAE=0.13), fits a lot of the random noise and thus, overfitting (Figure AP2.6)

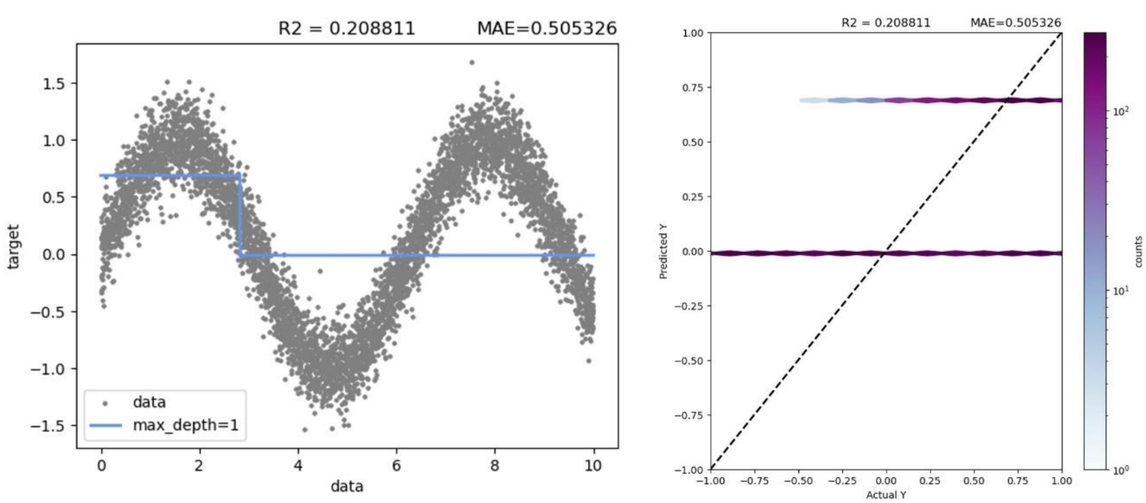

Figure AP2.2- Max depth =1 decision tree (light blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y. In this kind of scatterplot, the data density (color coded, white for low density and purple for high density) is shown within each one of the regularly spaced hexagonal bins over the 5000 datapoints. It is an interesting way to see regression results with such a high number of datapoints.

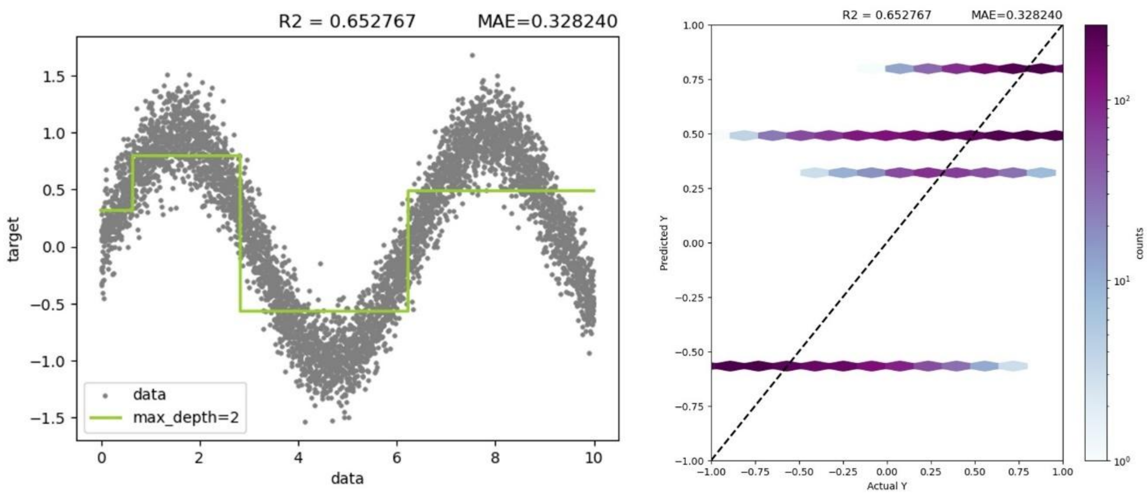

Figure AP2.3- Max depth =2 decision tree (mustard line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

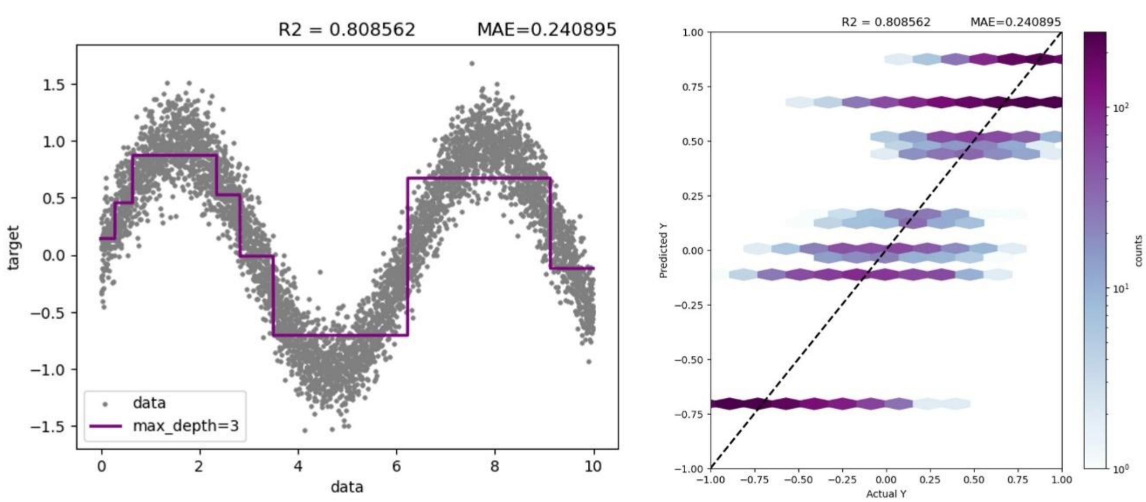

Figure AP2.4- Max depth =3 decision tree (mustard line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

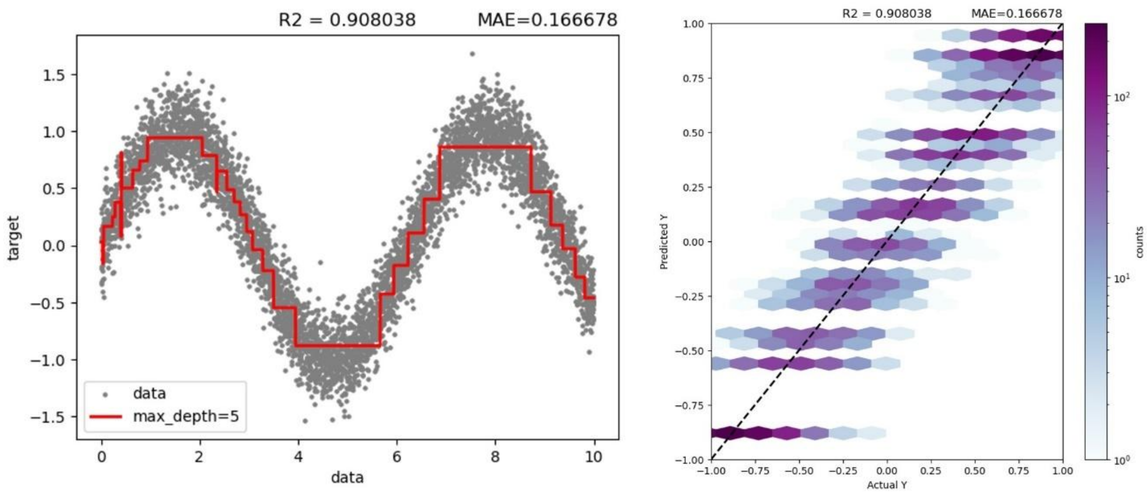

Figure AP2.5- Max depth =5 decision tree (red line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

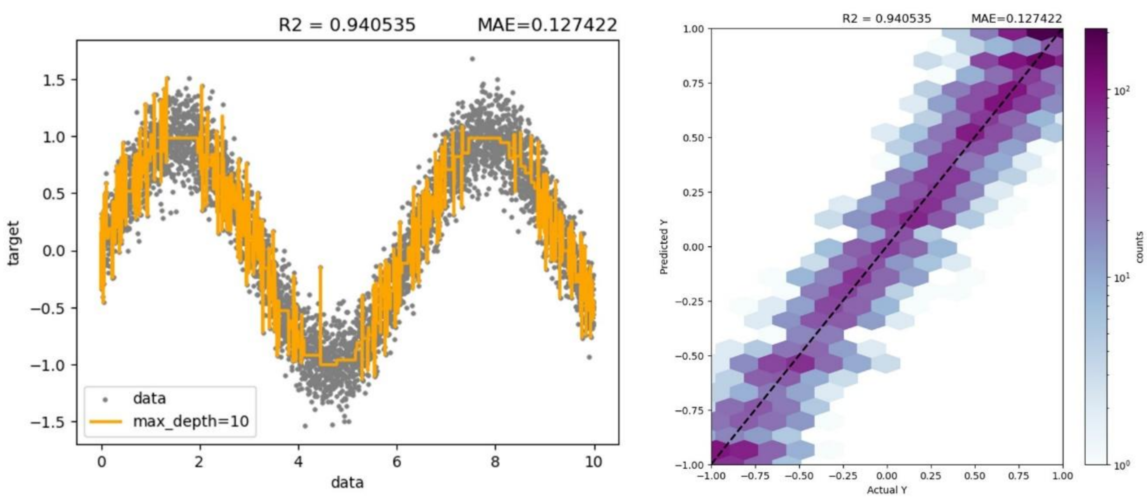

Figure AP2.6- Max depth =10 decision tree (orange line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

K-nearest neighbors (KNN)

KNN is one of the first machine learning methods that was invented by Fix and Hodges, (1951). In regression tasks using KNN, the prediction of the output parameter for a given data point in the training dataset is made by computing the average of the values of its K nearest neighbors, where K is the main hyperparameter. A secondary hyperparameter is a distance metric. In our case, we used Euclidean for the nearest neighbor's distance calculation. Figures AP2.7 to AP2.11 show different regression models using K values of 10, 50, 100, 1000, and 2000 nearest neighbors that are averaged for the prediction of y. As can be seen in Figure AP2.7, very low values of K (relative to the number of datapoints) are prone to overfit, whereas very high values tend to underfit data (Figure AP2.11). For this very simple nonlinear function, a KNN with K=50 (Figure AP2.8) has captured most data parameters (amplitude and wavelength of the sine function) and avoiding fitting noise.

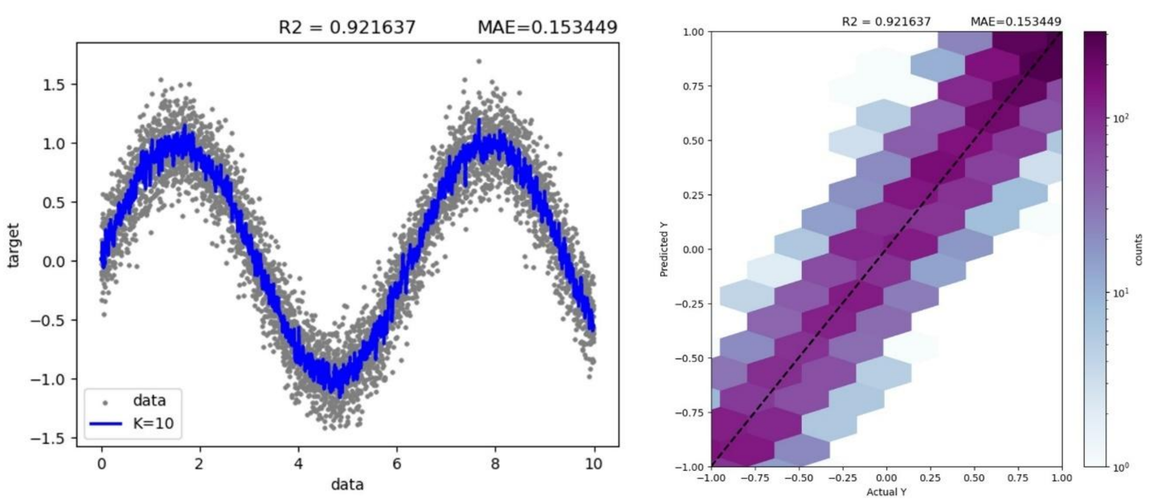

Figure AP2.7- K =10 KNN regression model (blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

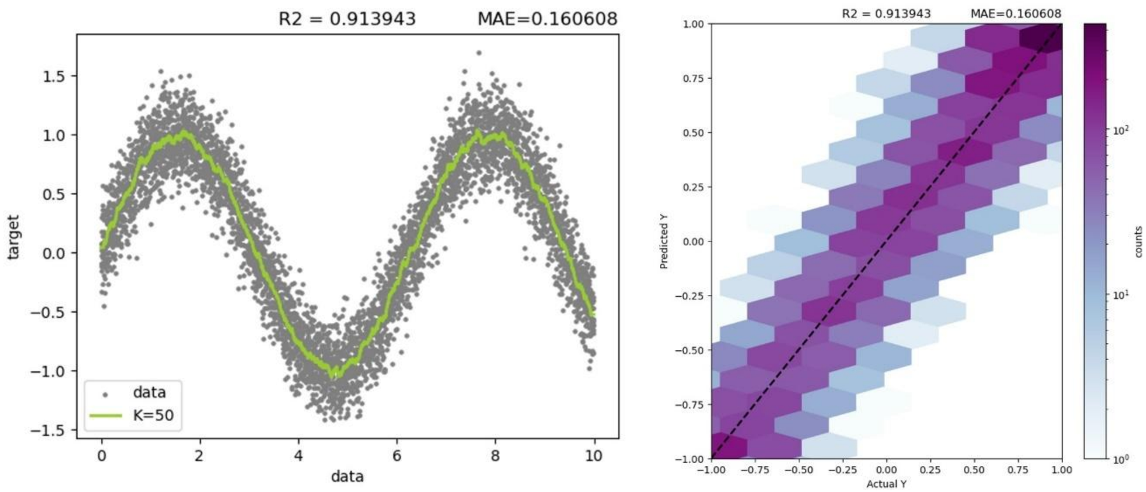

Figure AP2.8- K =50 KNN regression model (blue line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

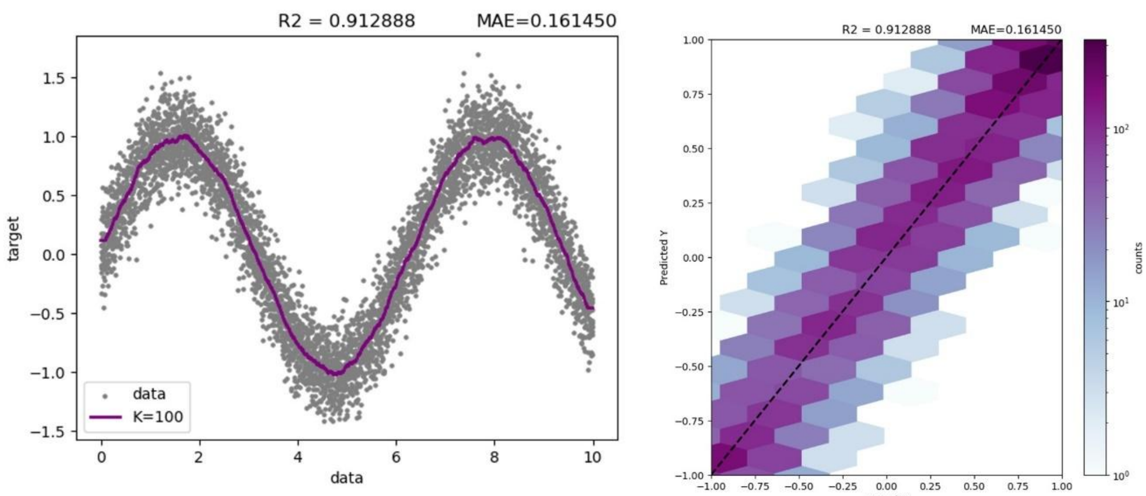

Figure AP2.9- K =100 KNN regression model (purple line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

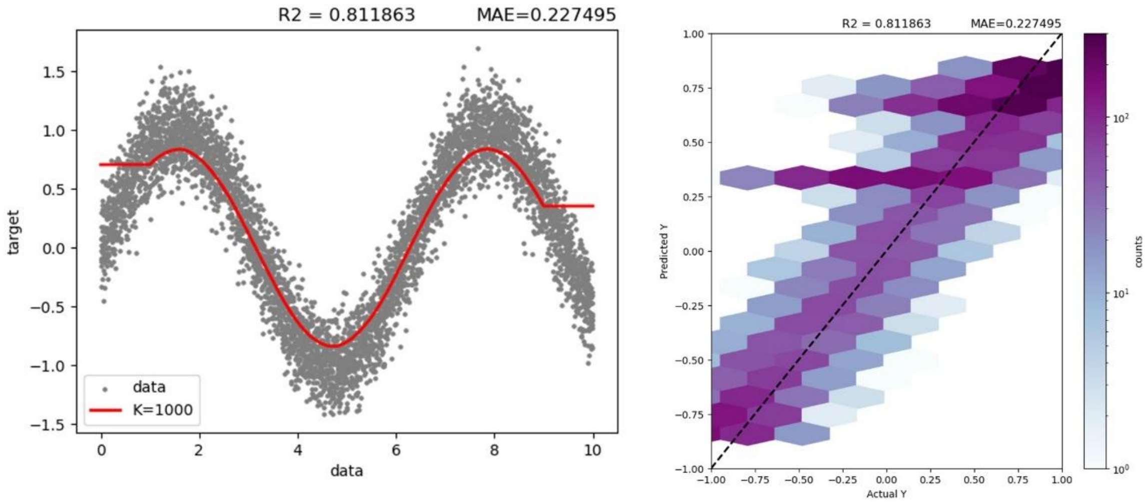

Figure AP2.10- K =1000 KNN regression model (red line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

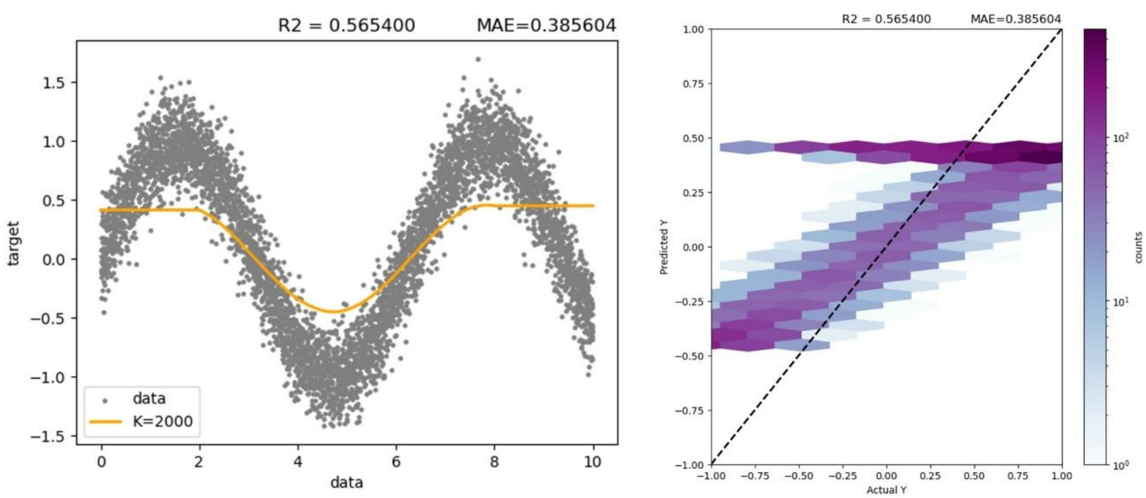

Figure AP2.11- K =2000 KNN regression model (orange line) fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

Support vector machines

Another very commonly used supervised classification (support vector classification) and regression (support vector regression – SVR) algorithm for multiple purposes, was invented by Vladimir Vapnik in 1964, but the first publications came only 30 years later in 1995 by one of his students (Cortes and Vapnik, 1995). It uses support vectors (datapoints) to build a hyperplane that fits the data given and error margin (epsilon tube). When working with non-linearly separable data (our toy example and majority of real-world problems) a kernel trick is applied by emulating a higher dimension space and, thus, making them linearly separable (Figure AP2.12)

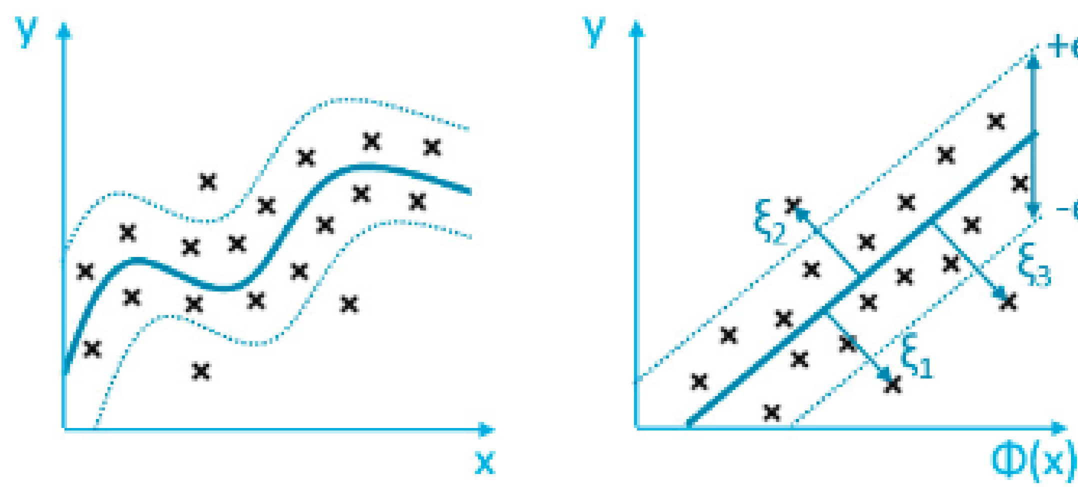

Figure AP2.12- Illustration of the kernel trick which was applied to linearize the data and then apply the SVR using the support vectors (ξ, blue arrows) and epsilon parameters (+ℇ, - ℇ), after Bock et al., (2020).

In our toy example, the main hyperparameter used was the epsilon value (allowed error margin). We started with SVR using epsilon=0.1 (Figure AP2.13) and then, tested epsilon values of 0.5, 0.8, and 1 (Figures AP2.14, AP2.15, and AP2.16, respectively).

It can be seen in Figure AP12.13 that epsilon=0.1 provided a very accurate prediction (R^2=0.92, MAE=0.16), without fitting noise (blue curve). As epsilon becomes higher, the predicting performance degrades and, after epsilon = 1 (Figure AP.12.16), it cannot reproduce the main sinusoidal function parameters of amplitude and wavelength.

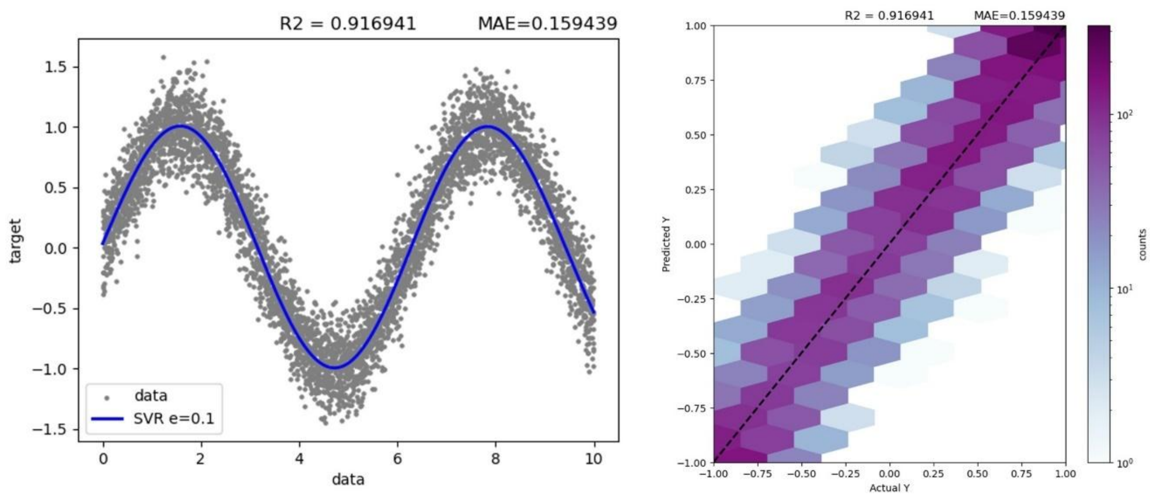

Figure AP2.13- SVR using epsilon= 0.1 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

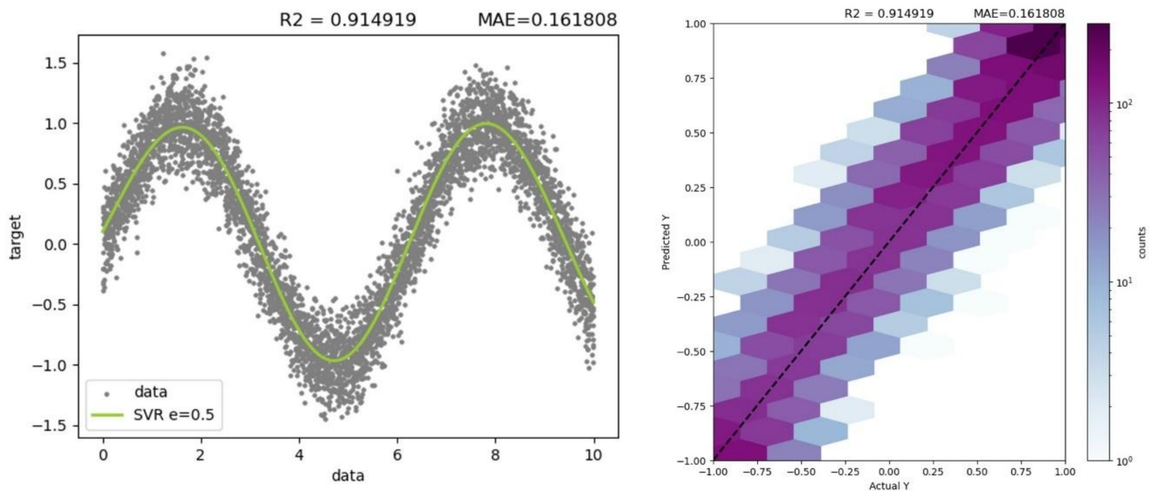

Figure AP2.14- SVR using epsilon= 0.5 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

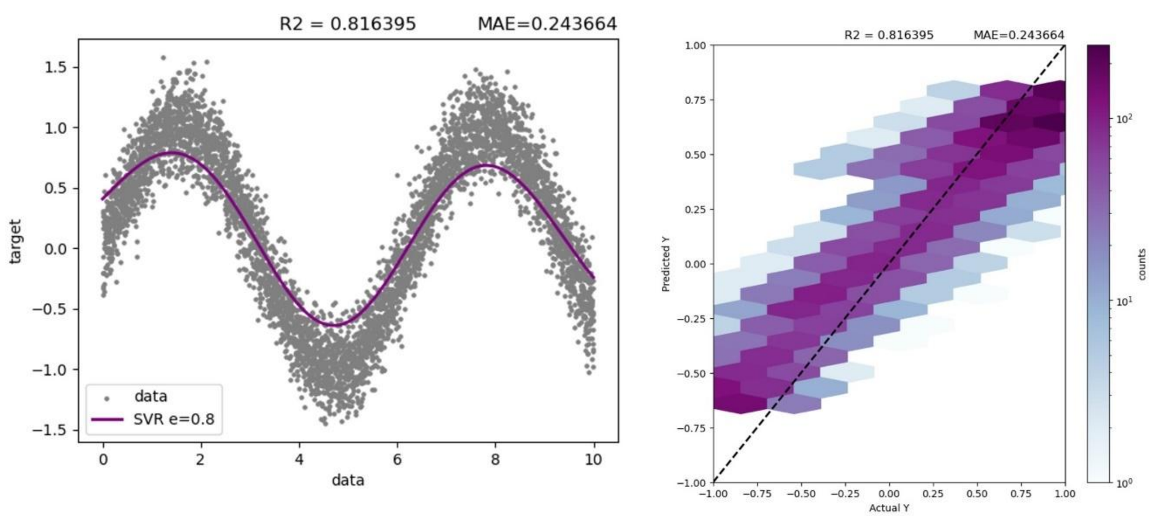

Figure AP2.15- SVR using epsilon= 0.8 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

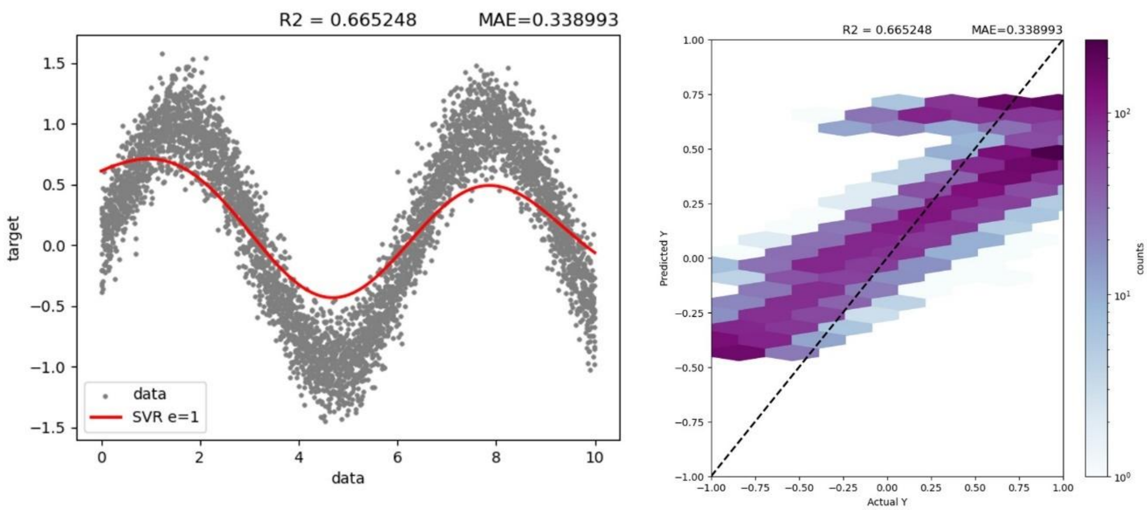

Figure AP2.16- SVR using epsilon= 1 fit to data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

Multilayer perceptron

Multilayer perceptron (MLP) is a kind of artificial neural network whose architecture mimics the human brain cells (neurons). It was first conceived by Rosenblatt (1958), however, due to technologic limitations, it was first implemented only twenty-eight years later by Rumelhart et al., (1986). The characteristic architecture of MLP is shown in Figure AP2.17, where a very simple neural network composed of one input layer (one independent variable x or feature), three hidden layers made of ten neurons each. In each neuron, there is an activation function that where the weights and biases (represented by the connections) are plugged into. These weights and biases are then optimized via descent gradient during backpropagation (Murphy, 2012). This architecture provided the best fit to data (R^2=0.90, MAE=0.17). All the layers are connected by weights and biases (represented by the connection lines in Figure AP2.17). At the hidden layers there are activation functions where the weights and biases values are plugged into. Those weights and biases are optimized by backpropagation with an objective function for error minimization. The main MLP hyperparameters used are the hidden layers size (3 hidden layers with 10 neurons each), the activation function within each neuron in the hidden layers (ReLU- rectified linear unit), the optimizer (Adam), and the regularization parameter (alpha= 0.0004). For this toy example, we tested different numbers (1, 2, 3, 5, and 10) of hidden layers with ten neurons each and fixed the other hyperparameters. The results are shown in Figures AP2.18 to AP2.22. The best fit was obtained with MLP regression using three hidden layers (R^2=0.90, MAE=0.17, Figure AP2.20), considering its simpler architecture when compared to the others with five and ten hidden layers which provided similar performance (Figures AP2.21 and AP2.22). The main take home point: too shallow neural networks may underfit your data, whereas too deep networks are prone to overfit your data.

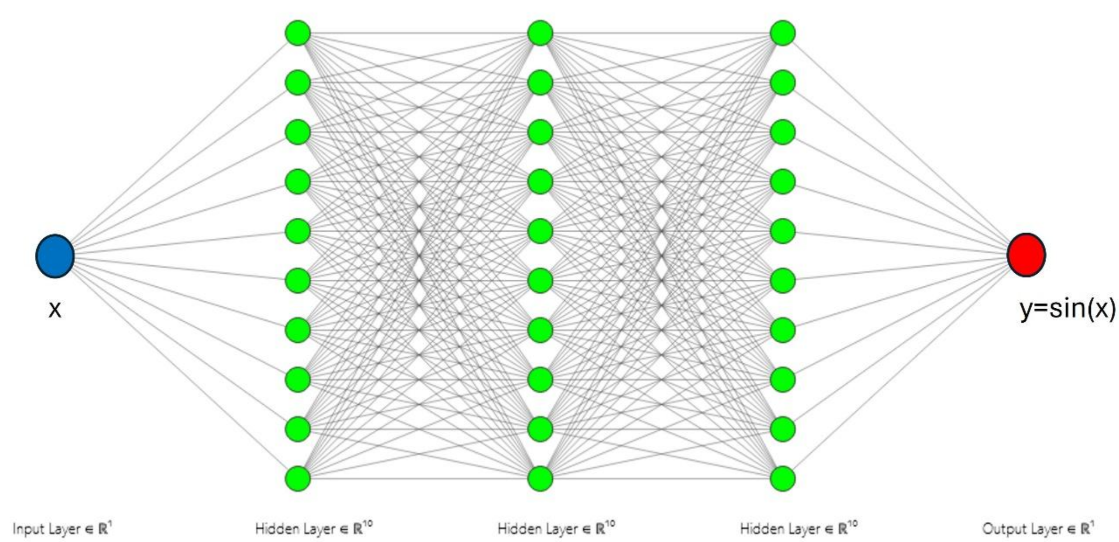

Figure AP2.17- MLP neural network architecture that best fits the data of this toy example. It is composed of one input layer of x values (5000 datapoints), 3 hidden layers with 10 neurons each, and an output layer with the prediction value that are closest possible to the analytic solution (y=sin(x)). For real world problems (like the ones described in this dissertation), more complex architecture is used such as: input layer with multiple features (or seismic attributes), up to dozens of hidden layers made of hundreds of neurons each. In the case of regression, the output layer is the single numerical prediction for each input layer datapoint.

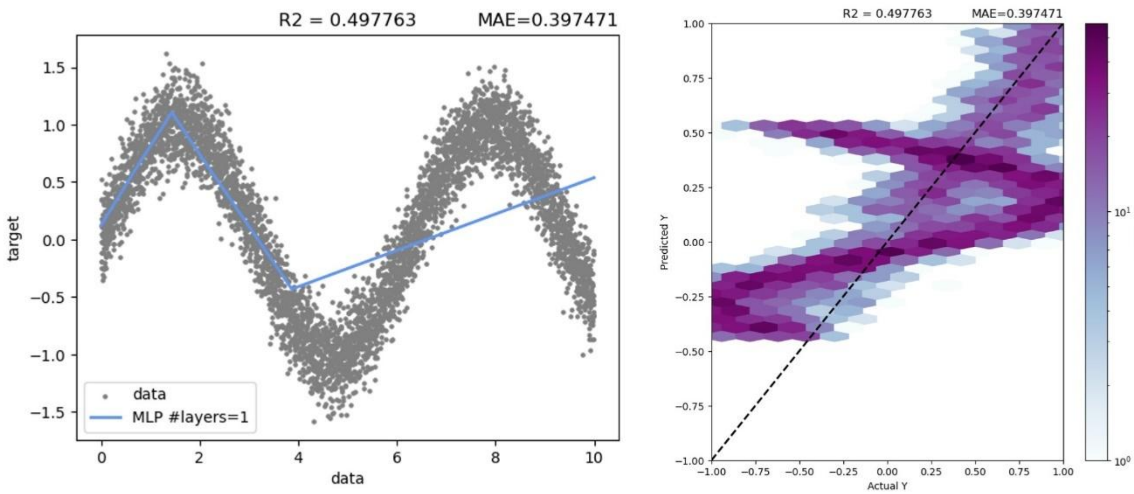

Figure AP2.18- MLP regression using one hidden layer with ten neurons to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

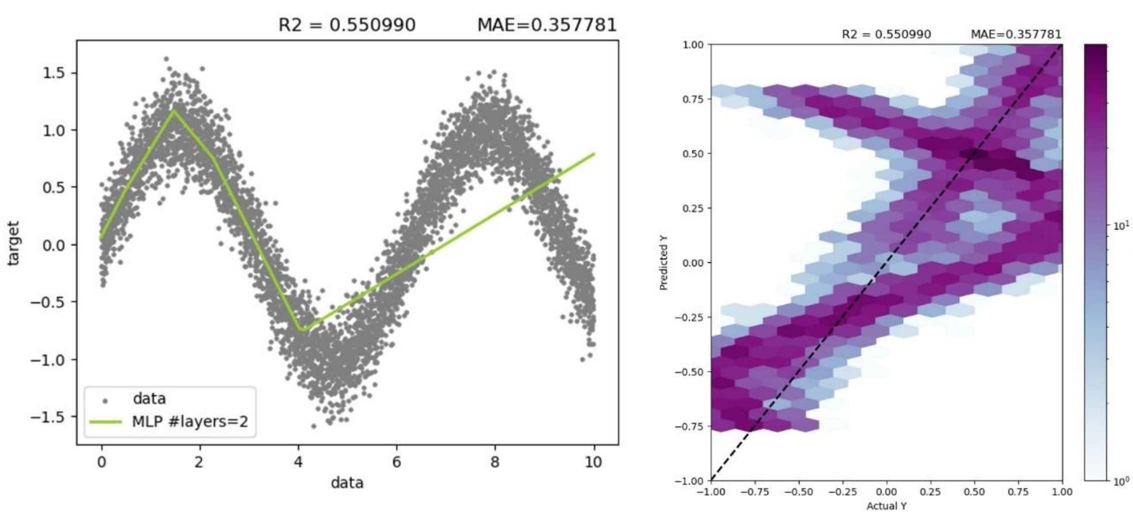

Figure AP2.19- MLP regression using two hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

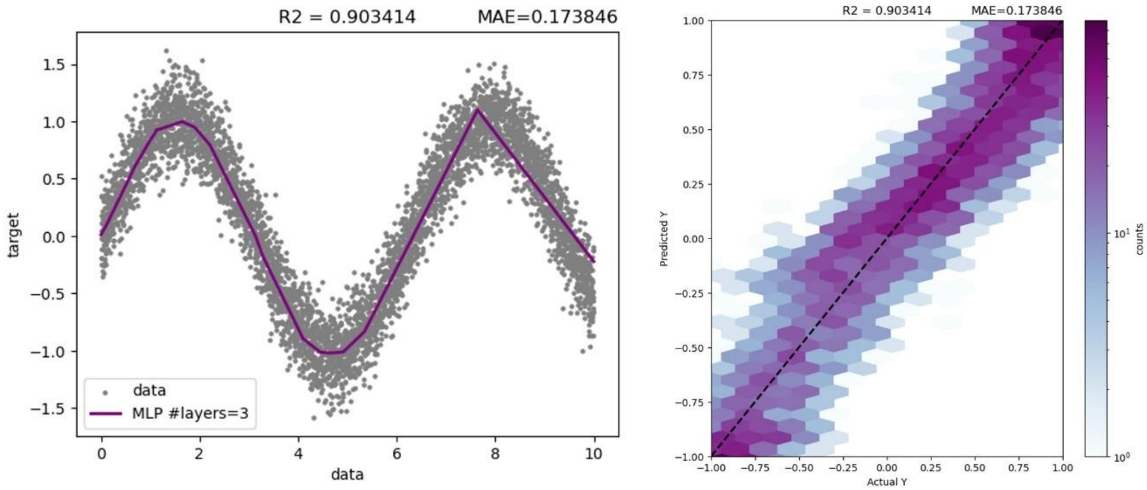

Figure AP2.20- MLP regression using three hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

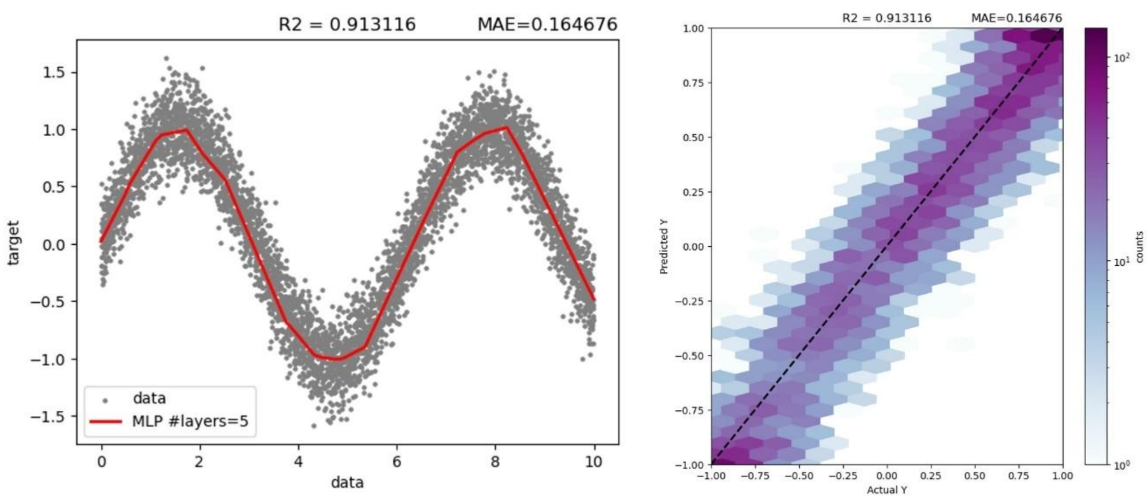

Figure AP2.21- MLP regression using five hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

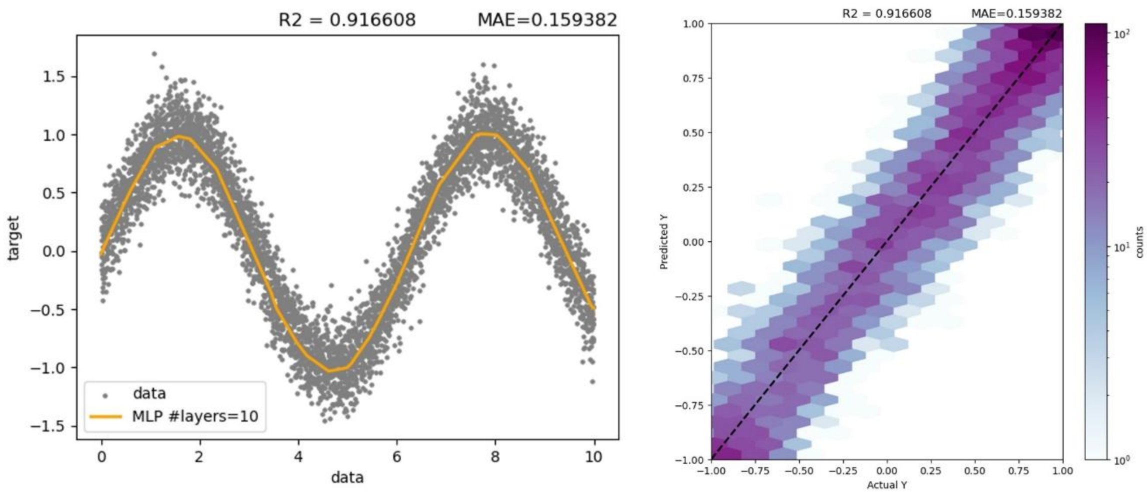

Figure AP2.22- MLP regression using ten hidden layers with ten neurons each to fit data (5000 datapoints, gray dots). On the right we have the hexagonal bin (hexbin) plot of actual versus predicted values of y.

References

Bock, F., Blaga, L.A., and Klusemann. 2020. Mechanical Performance Prediction for Friction Riveting Joints of Dissimilar Materials via Machine Learning. Procedia Manufacturing. 47. 615-622. 10.1016/j.promfg.2020.04.189.

Cortes, C. and Vapnik, V. 1995. Support-Vector Networks. Machine Learning, 20, 273-297. http://dx.doi.org/10.1007/BF00994018.

Fix, E. and Hodges, J.L. 1951. Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.

Ho, T.K. 1995. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada, pp. 278-282 vol.1, doi:10.1109/ICDAR.1995.598994.

Morgan, J.N. & Sonquist, J.A. 1963. Problems in the analysis of survey data, and a proposal. J. Amer. Statist. Assoc., 58, 415–434.

Murphy, K.P. 2012. Machine learning: a probabilistic perspective. MIT press. Cambridge, MA.

Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6):386.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning representations by back-propagating errors. Nature, 323:533–536.